Parallel worlds galore

The 50th anniversary of an astonishing scientific hypothesis deserves celebration. So too do the truly astounding tales of a literary genre that anticipated it.

From the evidence of our cover, you could be forgiven for thinking that you are holding a copy of *Nature* from an alternate universe. And if that were the way your imagination took off, it would be doing just what our cover seeks to do — celebrating the overlap between the world of science and the fables it inspires and feeds on. In particular, the 'Astounding Tale' of a plethora of alternate universes is at the same time a well-worn theme of science fiction and a valid, if speculative, way of understanding the ultimate implications of Schrödinger's wave equation.

The idea of a 'many-worlds' multiverse, introduced into physics 50 years ago this month by Hugh Everett, neatly highlights the intersection between science and science fiction — which is why our coverage of the anniversary spills from our News Features pages into our Books & Arts pages (see pages 15, 18, 23 and 25). For the most part, though, the two domains are themselves seen as alternates. It is a cliché of science popularization to proclaim that phenomenon X, once science fiction, is now science fact, as though the two were in some way mutually exclusive. This might suggest to some that science fiction is worthless; alternatively, it can tacitly imply that it is the job of science to reify the fancies of science fiction. Neither implication is useful.

The interaction between science and science fiction is more complex and symbiotic. Science fiction feeds on science. It also anticipates it. For good or ill, it articulates possibilities and fears: the notion of the super-weapon was commonplace in science fiction long before the Manhattan Project, and no debate about genetic technology seems complete without an appearance by Victor Frankenstein and his creature. More positively, science fiction provides crucial raw material — the minds of young people who will in time become scientists themselves. Not every science-fiction-reading teenager becomes a scientist, nor do all scientists grow up with shelves of Wells, Asimov and Le Guin by their beds. But the inspirational value is real.

This is not to say that science fiction is a childish thing, to be grown out of. But it does undeniably have a frequently childish character, one that reveals its true nature. Childhood is a time of games; games that allow their players' curiosity free expression while at the same time preparing them for a life in which every year brings novelties both anticipated and unlooked for. Science fiction, too, provides a way of exploring what is to come. Its main aim is not to foretell the future — indeed, the great Ray Bradbury once remarked that he wrote not to predict the future, but to prevent it. Yet even though it can be serious and frightening, it is not at heart a literature of warning, either. It is a literature of playfulness. Within the constraint of telling human stories about more-or-less human beings, it revels in the possibility of expanded physical and intellectual horizons.

> "Science fiction does not tell us what the future will bring, but helps us to understand what the future will feel like."

And above all it revels in the possibility of change. Serious science fiction takes science seriously, and its games provide a way of looking at the subjective implications of newly revealed objective truths of the Universe. Science fiction does not tell us what the future will bring, but at its best it helps us to understand what the future will feel like, and how we might feel when one way of looking at the world is overtaken by another.

To be sure, science fiction doesn't always connect in this way. It can be tired and cliché-ridden; the games it plays can be tedious, solipsistic power fantasies. And over recent years many of its finest practitioners have become so besotted by the endless new games that ever-accelerating progress allows them to play that their works can be inaccessible to the general reader. To demand that everything be accessible is to demand mediocrity — there is a role for dialogues that can be appreciated only by *cognoscenti*. But we believe that science fiction written for every scientist can be rewarding, too, which is why this issue sees the return of our popular showcase for short science-fiction stories, Futures (see page 104).

Science takes place in a cultural context. The many forward-looking, ever-changing worlds of science fiction provide one that is both fruitful and enjoyable. ∎

Enough talk already

Governments should act on researchers' attempts to engage the public over nanotechnology.

There hasn't been anything quite like it in the history of science. Over the past three years, in Europe, the United States and Australasia, a plethora of groups of scientists and other citizens have discussed nanotechnology in extended exercises in 'public engagement'. These initiatives have arisen partly because of the acute awareness by many that all is not as it should be either in the degree of trust in science and technologies on the part of the public, or in confidence in nanotechnology in particular.

As documented in a report published last week (see www.involve.org.uk/negreport), these initiatives have a number of common features. Definitions of the word 'nanotechnology' have been chewed over, optimism expressed that nanotechnology can benefit mankind, but — most importantly — concerns have been expressed about a lack of knowledge and regulation surrounding the impact of nanoparticles on health and the environment.

Equally notable were common outcomes for participants, where members of the public and scientists reported how much they had unexpectedly gained in understanding each others' perspectives.

Scientists appreciated how non-scientific participants were constructively interested and able to get to grips with key aspects of a complex subject. The latter, meanwhile, usually starting from zero, gained knowledge about what scientists are doing and about their motivation, and a greater awareness of the potential impacts — both real and hyped.

But what's the point of such engagement? One positive example inspired many subsequent activities: the 2004 report on nanoscience and nanotechnology by the British Royal Society and the Royal Academy of Engineering. This was productive not only in its content but also in the way the various processes of engagement enhanced its public credibility and helped ensure that the questions it addressed were of appropriate scope.

A more recent example was a series of 'Nanodialogues', workshops conducted by the UK think-tank Demos unveiled last week (see www.demos.co.uk/publications/nanodialogues). One workshop, involving Unilever, showed how issues of corporate social responsibility arise not only in manufacturing processes and products but also in corporate R&D. Another, conducted in a village near Harare in Zimbabwe, demonstrated how remote from reality are some claims made for nanotechnology's potential impact on water purification in the developing world.

A nanodialogue in Swindon involved engagement between 14 local citizens and scientists funded by two UK research councils based there. Here was an example of true 'upstream engagement', the idea — often suspiciously received by scientists — that both the public and scientists have something to gain from discussing future research prospects as an input to research funding. Most encouraging for researchers was the strong support by public participants for fundamental science.

A taste of true upstream thinking by nanoscientists can be found in blogs at http://ideasfactory.wordpress.com. These feature futuristic nanotechnology concepts, including software-controlled assembly of a variety of building-blocks (small molecules to nanoparticles) linked by covalent bonds, and the development of a flexible machine, computer language and compiler as an assembler of molecules and materials under atomically precise control.

These are potentially powerful enabling technologies. To take a view on their risks or values, either as a scientist or as a citizen, depends on imagined contexts of application. The benefit of the public engaging with scientists years or even decades ahead of the arrival of such technologies lies in the broadening of the bases of knowledge, mutual trust and — most importantly — critical appraisal. One challenge now is how to allow a much larger proportion of the public to share in those benefits.

> **"Most encouraging for researchers was the strong support by public participants for fundamental science."**

Few governments have put solid investment in the one type of research most consistently and urgently demanded by these groups — on the health and environmental risks of technologies already embedded in hundreds of products on store shelves. Commendably, a new Australian initiative in nanotechnology research includes such investment.

Regrettably, the governments of two countries that have taken strong leads in nanotechnology — the United States and Britain — have failed to respond. These governments and others not only need to act on this outcome of public engagement but must also integrate such processes into their departments' and agencies' activities. ∎

Discriminating on genes

The United States is belatedly establishing necessary protections in law. Others, take note.

In a rare departure, President George W. Bush, the House of Representatives and most of the US Senate have found a landmark piece of legislation on which they can agree. The Genetic Information Nondiscrimination Act of 2007 was passed overwhelmingly by the House late in April. The bill would bar insurers from using the results from genetic tests to deny coverage to new applicants, or from hiking the price of premiums for existing customers. It would also make it illegal for employers to use genetic information in hiring, firing or promotion decisions. (Shamefully, however, military personnel are excluded from the act's protections.) Such protections are sorely needed in many countries, but certainly in the United States, where employers and private insurers pay a significant proportion of health costs and have every reason to try to minimize the health risks of those they hire or insure.

Despite the endorsement of the White House and strong historical support for such a measure in the Senate, the bill is still awaiting a Senate vote. It has been delayed by a crammed legislative calendar and by behind-the-scenes wrangling of lawmakers over issues ranging from the measure's definition of a genetic test to whether the bill as written would allow violators to be sued twice for the same offence.

These negotiations should be concluded soon. There have seldom been so many good reasons to see a bill speedily enacted. Public concerns about the potential abuse of genetic information have risen steadily in the past decade. Leading voices, from the White House to top geneticists such as Francis Collins, argue that this threatens both the beneficial use of more than 1,100 clinically available genetic tests and the ability to conduct further research. Genetic studies rely, after all, on the good will of thousands of subjects who agree to have their DNA scrutinized. Without solid legal protection in place, people will remain justifiably wary of being tested, whether for research or for their personal health. The wealth of information now pouring out — for instance, in a raft of genome-wide association studies that are targeting suspect genes in complex diseases such as diabetes — will not be fully exploited, and the benefits of personalized medicine will be, at best, only partially realized.

Given the nature of its private-insurance system, it is apt that the United States should be leading the way with the current legislation. The Charter of Fundamental Rights of the European Union includes "genetic features" in its non-discrimination protections, and individual countries are charged with making their domestic laws compatible with it. Yet many countries — France and Austria are notable exceptions — have so far failed to enact an explicitly protective genetic information law. In this, they could take a leaf from the US book — provided that the book finally gets written. ∎

RESEARCH HIGHLIGHTS

Slow-speed sprinters

J. Roy. Soc. Interface doi:10.1098/rsif.2007.1095 (2007)

Elephants switch from walking to running at surprisingly low speeds, researchers have found.

John Hutchinson of the Royal Veterinary College in London and his colleagues previously provided some of the first evidence that elephants could run, clocking elephants in Thailand hitting top speeds of 25 kilometres per hour (J. R. Hutchinson *et al. Nature* **422,** 493–494; 2003). A follow-up study, in which elephants in UK safari parks were tagged with motion sensors, now reveals that the animals adopt a running gait once they hit 8 kilometres per hour. Above this speed, elephants use their back legs like 'pogo sticks' to drive their bodies forward over their stiffer forelimbs in a vaulting motion.

A. WOLFE/GETTY

COSMOLOGY

Gone without a trace

Nature Phys. doi:10.1038/nphys654 (2007)

'What happened before the Big Bang?' used to be a silly question, as time and space were thought to have begun there. But some theories of quantum gravity suggest that the Big Bang was a transition rather than a beginning, such that scientists might think about — and perhaps even observe — what went before.

Even so, says Martin Bojowald of Pennsylvania State University, limits will still apply to what we can learn. He shows, for a very simple quantum model of the Big Bang, that some of the conditions 'before' leave no discernible imprint on what comes 'after'. In particular, if the pre-Big Bang Universe was a mass of quantum fluctuations, essentially all information about these would have been lost in the Big Bang itself.

BIOCHEMISTRY

Game of tag

J. Am. Chem. Soc. doi:10.1021/ja070003c (2007)

Chemists have created a new fluorescent tag to label proteins and quantitatively monitor their interactions.

The tag belongs to a class of small molecules called biarsenical multi-use affinity probes (MAPs). These probes bind specific amino-acid sequences and are smaller than the fluorescent proteins commonly used as labels, making them less likely to interfere with protein function or localization.

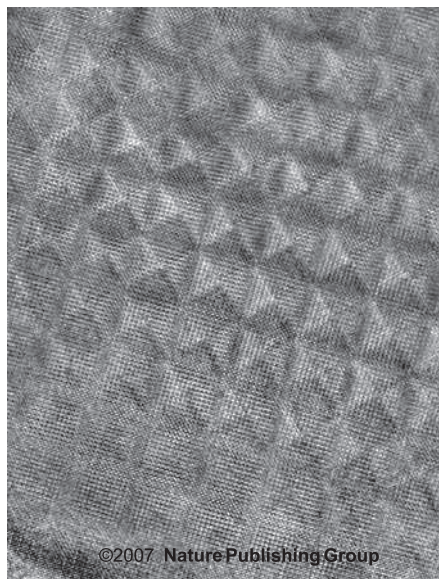Uljana Mayer and her colleagues at the Pacific Northwest National Laboratory in Richland, Washington, created a MAP that incorporates a red dye named Cy3 and recognizes a different amino-acid sequence to previously available MAPs. The individual probes can be used to simultaneously label different proteins, and interactions between two probes can be used to monitor protein–protein interactions.

MATERIALS SCIENCE

Squared to perfection

Nature Mater. doi:10.1038/nmat1953 (2007)

The geometric pattern shown below is an electron-microscope image of the surface of a crystal of a perovskite mineral. Its repeating structure has "extraordinary perfection", according to Beth Guiton and Peter Davies at the University of Pennsylvania in Philadelphia.

The pattern appears in a perovskite containing lithium and neodymium. The researchers propose that the material phase separates into square lithium-rich regions a few nanometres across, separated by lithium-poor stripes. Bumping up the ratio of lithium to neodymium tunes the pattern's periodicity, increasing the size of the squares.

The team suggests the oxide's surface could serve as a template for the assembly of nanoscale objects such as tubes or wires.

BIOTECHNOLOGY

Film ends badly

Proc. Natl Acad. Sci USA **104,** 11197–11202 (2007)

Bacteria hunkered down inside a biofilm can be targeted by bacteriophages — bacteria-destroying viruses — that have been armed with tools to break through the bacteria's biofilm defences, show Timothy Lu and Jim Collins of Boston University in Massachusetts.

Biofilms can help bacteria to thrive in places they're not wanted, such as in medical devices, by protecting the bacterial cells within a tough matrix of sugars, proteins and other components. These films are hard to remove, or to treat with antibiotics.

Lu and Collins genetically engineered a biofilm-busting enzyme into the genome of phage T7, which targets *Escherichia coli*. The enzyme breaks apart the biofilm matrix, so that the phage can attack the bacteria, hijacking their replication machinery for its own multiplication and then killing the bacterial cells. This offers a strategy for creating armies of phages tailor-made to attack biofilms of different bacterial pathogens.

SCIENCE SOURCE/SPL

MATERIALS SCIENCE

On Earth as it is in heaven

Appl. Phys. Lett. doi:10.1063/1.2752718 (2007)

There's no need to go into space to do microgravity experiments on crystal growth, say researchers. Applying a magnetic field to a sample can have the same effect.

Experiments have been flown on rockets and the International Space Station to study whether reduced gravity makes it easier to grow high-quality protein crystals. Quality is important when determining a protein's structure from its crystal form.

In a gravitational field, a crystal growing in solution creates a convection plume, because the buoyancy of the protein solution changes as the material precipitates out. This may affect the crystal's quality. Elias Vlieg of Radboud University Nijmegen in the Netherlands and his colleagues show that a magnetic-field gradient can suppress this potentially perturbing convection as effectively as microgravity, by acting on a protein's diamagnetism.
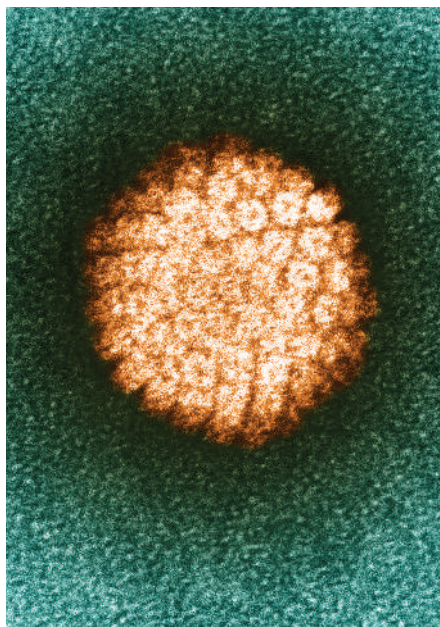
MEDICAL RESEARCH

Spermicide aids infection

Nature Med. doi:10.1038/nm1598 (2007)

Ingredients found in spermicides and vaginal lubricants can affect the transmission of genital human papillomavirus (HPV), the sexually transmitted infection to blame for almost all cases of cervical cancer, researchers have found.

John Schiller and his colleagues at the National Cancer Institute in Bethesda, Maryland, studied how HPV infects cells in a mouse model. The team used a pseudovirus (as human HPV is not readily available) tagged with fluorescent markers.

They found that the presence of the

spermicide nonoxynol-9 increased infection rate, possibly because the chemical damages the genital tract's inner layer. In contrast, carrageenan, a molecule found in many lubricants, inhibited infection, suggesting it might be used as a microbicide .

IMMUNOLOGY

Innate immunity to TB

J. Clin. Invest. **117,** 1988–1994 (2007)

A study of almost 200 people who had been exposed to tuberculosis (TB) has revealed a previously unappreciated facet of the immune system's defences against the disease. Researchers led by Adrian Martineau and Robert Wilkinson of Imperial College London found that TB immunity is boosted by high levels of neutrophils, a type of white blood cell.

These cells, which are part of the innate immune system, may be a more significant factor in TB resistance than the adaptive immune system — the target of vaccines such as the BCG.

The study shows that neutrophils produce proteins such as cathelicidin LL-37 and lipocalin 2, that defend against the TB bacterium *Mycobacterium tuberculosis*. Black Africans in the study had lower average neutrophil levels than south Asian or white volunteers — but as LL-37 production can be boosted by vitamin D supplements, this might offer another way to combat the disease.

NEUROSCIENCE

Dazed and confused

Nature Neurosci. doi:10.1038/nn1931 (2007)

Scientists at the University of Maryland, Baltimore, have identified a population of neurons that may contribute to impaired decision-making in drug addicts.

Thomas Stalnaker and his colleagues trained cocaine-addicted rats to associate different odours with either a reward or a punishment. When they reversed the associations, they found that cocaine-addicted animals were slower to adapt than healthy animals.

The researchers then monitored neural activity in a brain region involved in learning and conditioning, known as the basolateral amygdala, from implanted electrodes. This revealed a group of neurons that responded to the cues and that showed altered firing patterns in healthy animals — but not in cocaine-treated mice — when the cues were switched.

Inactivating the basolateral amygdala of addicted rats restored their ability to change their response to the cues.

JOURNAL CLUB

Allan Balmain
University of California,
San Francisco, USA

A cancer geneticist delves into family matters.

A mystery lies at the heart of a small family of growth signalling enzymes (K-Ras, H-Ras and N-Ras), which are widely mutated in human cancers. In culture, all three enzymes have similar functions, but different *ras* genes are associated with cancers in different tissues.

My laboratory, for instance, noted more than 25 years ago that skin cancers show activation of H-*ras*. Others have demonstrated that lung, colon and pancreatic cancers show activation of K-*ras*, whereas N-*ras* is the oncogene of choice in melanomas and some leukaemias.

What determines this intriguing specificity? Are the enzymes' functions somehow modified in certain tissues *in vivo*? Or is it regulation of the genes, affecting where and when they are expressed, that matters?

We may get some answers by

following the lead of an elegant study (N. Potenza *et al. EMBO Rep.* **6,** 432–437; 2005). In this work, the authors knocked out K-*ras* in mice, but simultaneously replaced the gene with its close relative H-*ras*, doctored to have the regulatory elements of K-*ras*. Mice can survive without the H-*ras* or N-*ras* genes (or even both of them) but usually die if K-*ras* is deleted. These mice, despite lacking K-*ras*, were viable and lived to a ripe old age.

This important observation provides novel opportunities to probe the mechanisms of cancer initiation. Are the mice lacking

K-*ras* now resistant to the lung and pancreatic cancers that are normally linked to K-*ras*? If yes, this would indicate a true requirement for the K-Ras protein in lung-cancer development; if not, the focus would switch to regulation.

A straw poll of Ras cognoscenti suggests that opinion is for now divided, but my group and others are working on this mouse model, and hope to have answers soon.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

The petaflop challenge

IBM's announcement on 26 June that it was about to smash the 'petaflop' speed barrier could herald a new era in computing. But unless the research and computing communities get their programming act together, they risk having few scientific applications that can take advantage of this huge increase in power, say experts.

Called Blue Gene/P, the first of the new high-powered computers should be operational next year, and IBM has already lined up potential customers in the United States, Germany and Britain. The largest planned configuration for the machines would run continually at 1,000 trillion floating point operations per second, or teraflops, and be capable of peak speeds of up to 3,000 teraflops (3 petaflops). That would make it between three and ten times faster than the machine that tops the latest TOP500 list of supercomputers, the IBM Blue Gene/L at Lawrence Livermore National Laboratory in California, which peaks at 360 teraflops.

"I'm very excited," says Ray Bair, head of the Argonne Leadership Computing Facility at the Argonne National Laboratory in Illinois, which has worked with IBM on developing the Blue Gene/P, and which will host a version of the hardware. "We're crossing a threshold," Bair adds, claiming that the increased power will at last allow researchers to build and run models in the way they have always wanted.

At prices of around $100 a gigaflop, peta-flop computers will start off in the $100-million range. A great deal of the added speed that money will buy simply comes from more processors. Supercomputer designers began taking parallel processing seriously in the 1990s, but few machines have been designed to work with more than 10,000 processors. In 2005, Blue Gene/L, which is almost three times as fast as its nearest competition, marked a significant step up with 131,072 processors. A 3-petaflop Blue Gene/P will boast 884,736 — a multitude that brings with it problems as well as promise.

The rapid recent growth in supercomputing power — Blue Gene/L is as powerful as the whole of 2002's TOP500 list combined — has come mostly from increases in the performance of the computer's component processors. But a few years ago that process "came to a grinding halt", says Jim Sexton, head of Blue Gene applications at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York. As the processors got faster and faster, they began suffering from disproportionate increases in power consumption and heat output.

To combat this, chipmakers made a historic switch. Since 2004, they have concentrated on increasing the number of processors on a chip, allowing the speed at which individual processors operate to plateau. So the dual and quad chips now common in laptops, offering two or four processors, may well be upgraded to 128 or 256 cores by 2015. This means that the cheap Linux supercomputing clusters common in universities would have hundreds of thousands of processors, and dedicated supercomputers might have hundreds of millions. The processors will not necessarily be blazingly fast — Blue Gene/P's 850-MHz chips are little faster than a Pentium III from 1999. But with their numerical advantage they won't have to be.

This new reliance on parallel processing for increased performance means that companies from Microsoft to Nintendo will have to rethink their software — and so will scientists. A few scientific applications fall into the favoured subset called 'embarrassingly parallel problems'. Genome analysis using BLAST software to compare sequences and mass spectroscopy for proteomics are generally fairly easy to parallelize, says Leroy Hood, president of the Institute for Systems Biology in Seattle, Washington. Each processor can take on a specific task without much reference to what all the others are doing. But other sorts of problem, in which many of the calculations depend on other calculations being done elsewhere, are not so tractable.

For the moderate levels of parallelism seen to date, it is possible to get by with the current practice of writing code, and designing models, in terms of linear sequences of instructions and then parallelizing once satisfied. To get the most out of massive parallel clusters and machines, that will no longer be an option. "Coding models running across as many as a million processors is a new challenge we have to meet," says Tim Palmer of the Europea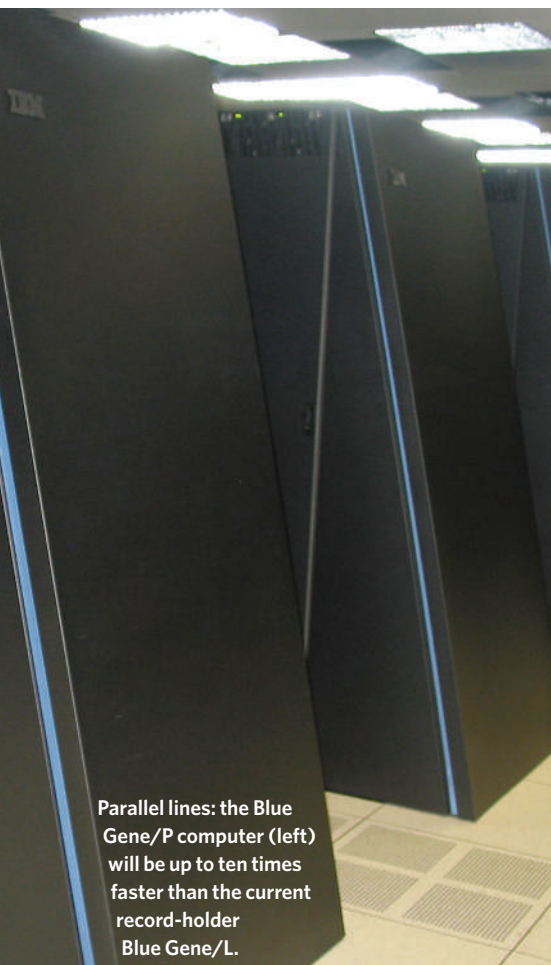n Centre for Medium-Range Weather Forecasts in Reading, UK, who is interested in petaflop machines for climate modelling. "We have no choice but to follow the hardware trends."

Scientists need to shift to thinking in parallel from the outset, designing hypotheses and code accordingly, says Horst Simon, associate lab director for computing at Lawrence Berkeley National Laboratory, California. He describes what is needed as nothing short of a "revolution in scientific programming". Such a revolution has been brewing for decades, but there hasn't been much storming of the barricades. "The high-performance computing community has been working on the parallel-programming problem for over 25 years. And frankly, we don't have much to show for it," wrote Intel researcher Timothy Mattson on his company's research blog last week. "On the software front, it's chaos."

> **"The scientific community is not very good at software development."**

MIND-ALTERING URINE
Alpha-male pheromones
cause females to make
brain cells.
www.nature.com/news

Parallel lines: the Blue
Gene/P computer (left)
will be up to ten times
faster than the current
record-holder
Blue Gene/L.

Such challenges, say experts, will expose serious weaknesses in the capacities of the scientific computation community. "The scientific community is not very good at software development," says Simon. He reckons his 20-year-old son, who writes gaming software for fun, is way ahead of most scientists in addressing the challenge of parallel programming.

"I'm amazed at what he can do just using open-source libraries," he says. Although there are exceptions, such as high-energy physics and bioinformatics, many labs keep their software development close to their chests, for fear that their competitors will put it to better use and get the credit for the academic application of the program. There is little incentive to get the software out there, says Simon, and such attitudes plague development.

One underlying reason is that in academia, software professionals are not given due recognition. "Why would you work at a low salary at a university, where the academic hotshot gets all the credit," asks Simon, "when you could work at a gaming company, be the hotshot and get stock options?" ∎
Declan Butler

UK science reshuffled

In the wide-ranging shake-up of government that accompanied Gordon Brown's arrival as Britain's new prime minister, science got both a new home and a new minister in charge. Although the move had not been announced at the time *Nature* went to press, the new minister will be Ian Pearson, previously Minister of State for Climate Change and Environment.

In a complex game of departmental division and recombination, Pearson will take office not in the Department of Trade and Industry (DTI), which used to be home to the science brief, but in the newly formed Department for Innovation, Universities and Skills (DIUS). This department marks a fusion between parts of the old DTI and part of the old Department for Education, and will also contain a new Office of the Chief Scientific Adviser. The trade part of the DTI's brief will now be shared by the Department for International Development and a new Department for Enterprise, Business and Regulatory Reform.

Pearson comes to the science brief having worked recently on launching a consultation process on the introduction of a cap-and-trade carbon-reduction scheme for up to 5,000 UK businesses. Welcoming the appointment, a spokesperson for the Royal Society said: "The government's track record in climate change has been good recently and Ian Pearson has obviously played his part in that. If he brings that track record to science we'd be very much encouraged." A strong emphasis on the environment at high levels in government seems assured by the appointment of David Miliband as foreign secretary; as Secretary of State for the Environment, Miliband was a strong voice on climate change.

Yet Phil Willis, chairman of the parliamentary science and technology select committee, expressed concerns that environmental science is about to hijack science as a whole. "A danger of the environment becoming so dominant," he told *Nature*, "is that other branches of science will be playing second fiddle." To do justice to science as a whole, Willis said, Pearson will have to "get to grips with a portfolio that is absolutely massive".

The creation of the DIUS, which in institutional terms will treat science as more closely linked to education than to enterprise, has received mixed reviews. "It makes sense to put science and higher education together," says Colin Blakemore, shortly to step down as chief executive of the Medical Research Council.

But David Brown, chief executive of the Institution of Chemical Engineers, warns that "Innovation and science should be located in the same ministry as enterprise and skills. I fear we will continue to be held back by the traditional interdepartmental boundaries." ∎
Katharine Sanderson

M. LIMA/AFP/GETTY

Climate for change: Ian Pearson will take on the science brief.

No solar hiding place for greenhouse sceptics

A study has confirmed that there are no grounds to blame the Sun for recent global warming. The analysis shows that global warming since 1985 has been caused neither by an increase in solar radiation nor by a decrease in the flux of galactic cosmic rays (M. Lockwood and C. Fröhlich *Proc. R. Soc. A* doi:10.1098/rspa.2007.1880; 2007). Some researchers had suggested that the latter might influence global warming through an involvement in cloud formation.

"This paper is the final nail in the coffin for people who would like to make the Sun responsible for present global warming," says Stefan Rahmstorf, a climate scientist at the Potsdam Institute for Climate Impact Research in Germany.

Claims that the Sun, rather than raised levels of greenhouse gases, has been responsible for recent warming have persisted in a small number of scientists and in parts of the media. Mike Lockwood, a physicist at the Rutherford Appleton Laboratory in Chilton, UK, says he was "galvanized" to carry out the comprehensive study by misleading media reports. He cites 'The Great Global Warming Swindle', a television programme shown in March by Britain's Channel 4, as a prime example.

Together with Claus Fröhlich of the World Radiation Center in Davos, Switzerland, Lockwood brought together solar data for the past 100 years. The two researchers averaged out the 11-year solar cycles and looked for correlation between solar variation and global mean temperatures. Solar activity peaked between 1985 and 1987. Since then, trends in solar irradiance, sunspot number and cosmic-ray intensity have all been in the opposite direction to that required to explain global warming.

In 1997, Henrik Svensmark, a physicist at the Danish National Space Center in Copenhagen, suggested that cosmic rays facilitate cloud formation by seeding the atmosphere with trails of ions that can help water droplets form (H. Svensmark and E. J. Friis-Christensen *J. Atmos. Solar-Terrest. Phys.* **59,** 1225–1232; 1997). He proposed that, as a result of this, changes in the Sun's magnetic field that influence the flux of cosmic rays could affect Earth's climate. This led to claims that cosmic rays are the main influence on modern climate change.

Even in the face of the new analysis, Svensmark insists that solar theories should not be dismissed. "If you look at temperatures in the troposphere, there is a remarkable correlation with solar activity," he says. Lockwood insists that none of the tropospheric data show the trend that the solar theory would need.
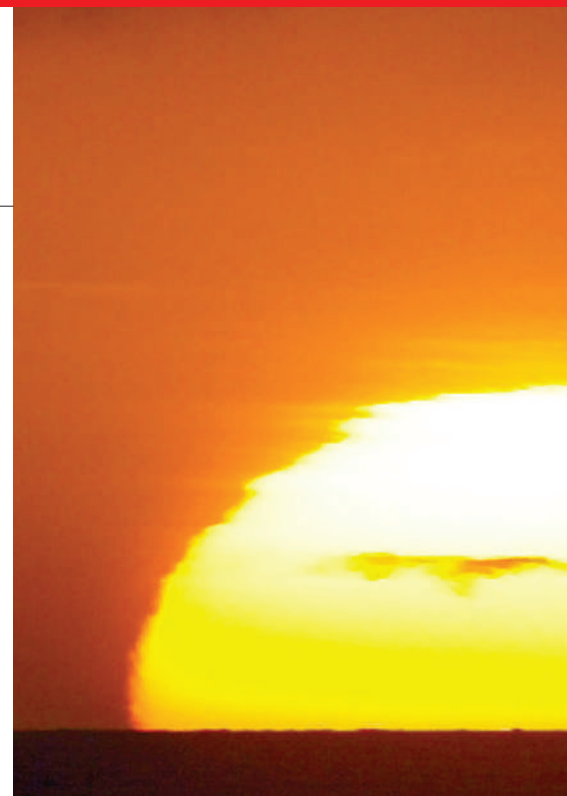
Nir Shaviv, an astrophysicist at the Hebrew University of Jerusalem who has championed a Sun–climate link, argues that there may be a lag in Earth's reaction to the Sun because of the thermal inertia of the oceans.

But other climate researchers find the idea of a 'hidden' time lag unconvincing. "With each year, and with each new set of data that comes in, a time lag becomes ever more unlikely," says Urs Neu, deputy head of ProClim-, the climate and global change forum of the Swiss Academy of Sciences in Bern.

On other timescales, however, Sun–climate links may remain worthy of study. "Climate change is a cocktail of many effects," says Jasper Kirkby, a physicist at CERN, the European

Hot topic: solar activity peaked too early to have caused the current trends in climate change.

Cosmic-ray results auger well for future

The first substantial results from the world's largest detector of cosmic rays are being presented this week at a conference in Mexico.

The Pierre Auger Cosmic Ray Observatory is already the largest cosmic-ray observatory in the world even though it is only three-quarters finished. It sprawls across more than 3,000 square kilometres of the Argentine pampas near the city of Malargüe — an area a tenth the size of Belgium. Researchers from the 15 countries collaborating on the US$50-million project hope that after its 1,600

A small component of a large observatory.

detectors and 24 telescopes have been commissioned around the end of this year the observatory may reveal the origins of the highest-energy particles bombarding Earth from space.

Among the results being presented at the International Cosmic Ray Conference in Merída, Yucatán on 3–11 July are new measurements of the energy spectrum of cosmic rays. The team confirms that above a threshold energy known as the GZK cut-off the number of cosmic-ray events falls sharply. Particles that start off with energies above this threshold are expected to be slowed down by interactions with the cosmic microwave background as they travel. Cosmic rays significantly above the cut-off energies would thus have to come from relatively nearby, which would imply that they were produced by strange new physics, such as the spontaneous decay of unknown massive particles.

An experiment in Japan called the Akeno Giant Air Shower Array sparked excitement when it failed to see the cut-off in data obtained up to 2004. Since then, though, data from the High Resolution Fly's Eye — a detector in Utah — have

**ASTEROID MISSION
GEARS UP FOR LAUNCH**
NASA mission leaders hope
Dawn will rise on time.
www.nature.com/news

The great gig in the South

When Al Gore and his colleagues dreamed up the idea of a climate-change consciousness-raising Live Earth event featuring concerts on every continent, they came up against a problem. The event's 7 July date corresponds with midwinter in the Antarctic, when no one gets on or off the seventh continent. So flying in Bono and Madonna to frolic with penguins would not be an option.

Enter Nunatak, a five-strong band of guitar, bass, drums, fiddle and sax, whose members are part of the crew at the Rothera Research Station, run by the British Antarctic Survey (BAS).

It is not the only band on the continent. Every New Year's Day, the US McMurdo base hosts the Icestock festival, and even in winter there's plenty of music. "All the bases have bands except maybe Bird Island, as there are only four overwintering people there," explains Julius Rix at the BAS's more southerly Halley base. But being the farthest north of the bigger bases gives Nunatak an advantage. "Rothera has double the satellite bandwidth compared to us," says Rix. And so Nunatak got the gig.

Even so, the webcam set-up on Rothera isn't quite up to sending broadcast-quality sound and visuals live to the world. "To do it live we would need a very fast satellite link," says Matt Balmer, an electronics engineer and Nunatak's guitarist and songwriter. So

> "They will be throwing their underwear."

the performance will be recorded 'as live' a day or so beforehand. "We're effectively e-mailing it in," says Balmer.

Despite the technical delay, the performance should be pretty authentic. The band plans to perform outside — which might make their instruments "sound a bit funny", according to the BAS's Linda Capper, back in balmy Cambridge, UK, not withstanding the rigours faced by the performers.

Balmer is tough enough to dismiss such worries: "It's only minus six outside; you get used to these sorts of temperatures."

Nor does the prospect of a global audience chill him, in part because it doesn't feel that real. "We don't have an appreciation of it," says Balmer. "We see the world through the Internet." The last ship out of Rothera left in April. "Until October we won't see anybody else."

Saxophonist and marine biologist Alison Massey expects the rest of the base to be an appreciative audience, not least because after weeks of rehearsal at close quarters, "it will be the last time they'll hear the songs". Balmer hopes for something a bit more rock and roll: "There will probably be groupies throwing their underwear at us," he says. With only 17 people in the audience, though, matching the knickers to the punters may not be too hard.
Katharine Sanderson

particle-physics laboratory near Geneva in Switzerland, who is leading an experiment aimed at simulating the effect of cosmic rays on clouds. "Past climate changes have clearly been associated with solar activity. Even if this is not the case now, it is still important to understand how solar variability affects climate."

Ken Carslaw, an atmospheric scientist at the University of Leeds, UK, points out that solar effects might still be possible. They might have acted to cool the climate in recent decades, but been overwhelmed. If so, the climate could be more sensitive to greenhouse gases than is generally thought, and future temperature increases might be greater than expected if a countervailing solar effect comes to an end. ■
Quirin Schiermeier

provided evidence for the reality of the cut-off. The new Pierre Auger data seem to confirm that the cut-off is real.

If there is no new physics in the form of particle decays or other exotic effects, though, there should be plenty of new astrophysics in understanding what sort of celestial mechanism might be pumping up incoming protons to energies millions of times higher than those produced by the most powerful Earth-based accelerators. One crucial step is to identify where in the sky they are coming from. Alan Watson of the University of Leeds, UK, a spokesman for the Pierre Auger project, says that the team has identified some interesting candidate regions of the sky, but won't reveal the details until it has obtained more data. "Unless I talk in my sleep, even my wife doesn't know what these regions are," he says. ■
Jenny Hogan

Antarctic brass monkeys: Nunatak will perform in sub-zero temperatures for Live Earth.

W. BURGESS/REUTERS

BRITISH ANTARCTIC SURVEY

SCORECARD

▲ **San Francisco's offices** Mayor Gavin Newsom has banned bottled water for thirsty city workers, who will now be offered only environmentally friendly water from the tap.

▼ **TV remote controls** Even the humble remote might soon feel too much like hard work. Japanese firm Hitachi has unveiled a 'brain interface' that it hopes will ultimately allow users to change channels just by thinking about it.

ON THE RECORD

"The conservation of the bald eagle is a true reflection of the concern Americans have for the environment."

George Fenwick, president of the American Bird Conservancy, is pleased by the news that the nation's emblem is to be removed from the Endangered Species Act's list of threatened species.

"It's the same as you feel after a plate of spaghetti."

Alessandro Sannino, Italian researcher and inventor of the 'fill pill', explains how his invention, which is made out of the same absorbent material as nappies, helps dieters by expanding to take up space in the stomach.

OVERHYPED

The iPhone
Yes it's beautiful. Yes it's multifunctional. But ultimately it's only a telephone, and therefore not worth camping out for days on end to get your hands on.

Sources: Reuters, Fox News, American Bird Conservancy, Daily Mirror, Associated Press

SIDELINES

APPLE

C. RUSS, BROAD INST.

Sequencing machines such as the 1G are revolutionizing the speed at which DNA is analysed.

Faster still and faster

A new generation of sequencing machines is broadening horizons for users. Various groups have recently performed epigenetic studies — looking at modifications to the genome that control its expression — that would have been utterly impractical using old technologies.

The latest approach fishes out all the DNA associated with a given marker, such as one of the histone proteins used to package genes in chromosomes. Then, instead of comparing each piece of DNA with a library of previously isolated sequences, as used to be done, scientists simply sequence the whole lot.

The key to this approach is new technology such as that sold by Solexa, a company that earlier this year merged with Illumina of San Diego. "The amount of DNA sequence being produced by these machines is staggering," says Steven Jones, associate director of the Genome Sciences Centre at the British Columbia Cancer Agency in Vancouver. Working flat out, a Solexa 1G machine could triple the total amount of DNA sequence contained in the GenBank database in just one year.

Jones's group looked at histone changes that control which regions of DNA can be read[1]. Meanwhile, scientists from the Broad Institute in Cambridge, Massachusetts, and Massachusetts General Hospital in Boston used a Solexa machine to examine two types of histone modification in mouse cells. Their paper[2], published online in *Nature* on 1 July, describes how these modifications change during development, and how such changes can either keep cells poised to switch fates, or close down their future options.

"The excitement about this paper is that we now have a means of studying cellular state in a high-throughput manner," says Bradley Bernstein, a pathologist at Massachusetts General Hospital who co-led the work with Broad director Eric Lander. "We certainly couldn't have done this on a genome-wide scale before." (See 'Sequencing revolution ushers in new era'.)

This advance is already unveiling new biology, says Keji Zhao of the US National Heart, Lung and Blood Institute in Bethesda, Maryland. Zhao's group has used it to decipher the messages encoded by two types of epigenetic mark produced by adding methyl groups to DNA[3]. Zhao says scientists' ability to take large-scale, complete snapshots means it might one day be possible to catalogue all the non-genetic alterations that control how genes are expressed in various cells and at all stages of development. And progress could be breakneck: Zhao's group took delivery of its first Solexa machine in January and published its results in May.

Beyond epigenetics, says Lander, other ambitious projects beckon. For example, this May, the genome of genomics pioneer James Watson was bared to the world[4] after being sequenced using technology from 454 Life Sciences, a company based in Branford, Connecticut, that has

just been acquired by Roche Diagnostics. This got institutions such as the US National Human Genome Research Institute thinking seriously about sequencing the complete genomes of hundreds or even thousands of people. The technology might also be used to identify new organisms, such as obscure bacteria living in complex microbial communities.

But Lander cautions that all this will take a lot of work as scientists map out how best to use the latest advances. The new technologies typically read out much shorter stretches of DNA than the older generation of sequencers did. And new methods must be developed for preparing samples and assembling these short reads into whole genomes. "It's going to be a nightmare for a year or two, as we try to fit old or important applications to these strange new platforms," Lander says. "But the return is going to be tremendous." ■

Erika Check

1. Robertson, G. *et al. Nature Methods* doi:10.1038/nmeth1068 (2007).
2. Mikkelsen, T. S. *et al. Nature* doi:10.1038/nature06008 (2007).
3. Barski, A. *et al. Cell* **129,** 823–837 (2007).
4. *Nature* **447,** 358–359 (2007).

Sequencing revolution ushers in new era

The genome sequencing centre at the Broad Institute in Cambridge, Massachusetts, has long been what ecologists call a monoculture: rows of identical sequencing machines, namely the trusty Applied Biosystems 3730s on which much of the human genome was sequenced. But all that is beginning to change.

First came two new types of machine, one made by 454 Life Sciences and the other by Solexa (see main story). Both use modified versions of the 'Sanger' sequencing method used by the old ABI workhorses, but add new wrinkles that allow vast numbers of short fragments to be sequenced in parallel.

In June, the Broad Institute and several other major sequencing centres received the first shipments of a new ABI sequencer that uses a sequencing chemistry that is fundamentally different from that of Solexa and 454. These machines read the cutting of DNA strands as opposed to their synthesis when sequencing. This technique requires serious computing power, as it produces terabytes of data at a time.

What all three instruments have in common is speed: Both the Solexa and ABI technologies claim to be able to identify 1 billion bases of sequence in a single run, taking just days to produce the equivalent of one-third of the human genome. 454 technology produces runs of about half that size.

Chad Nusbaum, co-director of the Broad's sequencing programme, says this is just the start. Other companies are developing even faster technologies that they're not discussing publicly yet. The new entries are changing the way sequences are harvested at the Broad and other large sequencing centres, such as those at Washington University in St Louis, Missouri, and the Wellcome Trust Sanger Institute in Hinxton, Cambridge, UK.

"This is a very aggressive horse race," Nusbaum says. "Companies are working on new techniques that in two or five years might dramatically rewrite the state of the technology, and that's a great thing for science." **E.C.**

War of words erupts over fossil dig

The Olduvai Gorge, perhaps the most famous location in palaeoanthropology, has become a battleground for two research groups. Last week the rivals met to resolve their differences, but tensions remain high.

The Tanzanian site in question is where, in 1959, Mary Leakey found the first skull of the species *Paranthropus boisei*. The following year, the Leakey team returned to find the first fossils of *Homo habilis*. The discoveries at Olduvai encouraged researchers to follow the course of Africa's Great Rift Valley up through Kenya and Ethiopia, where a host of new human relations were found in subsequent decades.

Numerous teams have tried to pin down the relationships between the bones and artefacts, and to piece together a sense of the lives the various species led two million years ago. For 18 years, the Olduvai Landscape Paleoanthropology Project (OLAPP) — led by anthropologist Robert Blumenschine of Rutgers University in New Brunswick, New Jersey, archaeologist Fidelis Masao of the University of Dar es Salaam and Jackson Njau, principal curator at Tanzania's National Natural History Museum in Arusha — has collected plant and animal specimens to learn how these early relatives of man lived in the region (R. J. Blumenschine *et al. Science* **299,** 1217–1221; 2003).

Last summer, the OLAPP team was distressed to learn that Tanzanian officials had issued permits to a group led by Manuel Domínguez-Rodrigo, of Complutense University in Madrid, and Audauz Mabulla, of the University of Dar es Salaam, to dig within the OLAPP region. The OLAPP researchers then found the competing group a kilometre away from their campsite, probing trenches the OLAPP team had dug near the bed where Leakey uncovered 'Zinj', the original *P. boisei* skull.

In general, researchers like to collect material in a standardized way. If a new group disturbs a site, it can harm the previous team's study if, for example, different collection techniques are used. The OLAPP team objected to the Tanzanian authorities and the Domínguez-Rodrigo/Mabulla excavation was stopped.

This prompted an angry response from Domínguez-Rodrigo and Mabulla on the website of the Cultural Association for the Study of the Quaternary in the Cantabrian Region (www.accuca.conectia.es). Domínguez-Rodrigo also attacked the OLAPP group in newspaper articles in Madrid and Dar es Salaam. On the website, Domínguez-Rodrigo and Mabulla

The first remains of *Paranthropus boisei* and *Homo habilis* were found at the Olduvai Gorge in Tanzania.

doubled OLAPP's grant total (US$2 million instead of $1 million), said that OLAPP had published only two articles (a project bibliography notes 32 articles, 8 book chapters and 12 graduate theses) and accused OLAPP leaders of "unethical behaviour" by blocking them using "political influence".

The OLAPP researchers summarized the charges in their abstract for the first East African Association of Palaeoanthropology and Palaeontology Conference, which leading researchers attended on 18–20 June in Nairobi, Kenya. In their abstract, the OLAPP researchers wrote that Domínguez-Rodrigo falsely accused them of "bribing Tanzanian officials" and characterized the competing group's approach as "antagonistic, unethical, uncollegial and counterproductive".

But the OLAPP leaders did not attend the meeting, instead rushing into the field in June to monitor the site as the competing group arrived, having received new permits. The OLAPP group sent a letter to the Nairobi meeting on 17 June, apologizing for not attending and petitioning attendees to "renounce" the competing group and to support the OLAPP's call for "an apology and retraction of allega-

> **"The situation is awkward for both groups."**

tions". But the meeting's organizers decided that in the interests of fairness the letter would not be read to attendees.

In a phone interview from Olduvai last week, Domínquez-Rodrigo said an apology wasn't necessary. There was no specific accusation of "bribery", he said, but of "unethical behaviour and interfering in local politics". He declined to elaborate on these charges so as not to interfere with ongoing negotiations.

"We've taken all the measures we could to avoid conflict," he said. "The situation is awkward for both groups." Last week the two teams agreed to cohabit for the season to come.

Domínguez-Rodrigo says he wants "to enlarge the area where Leakey excavated", applying modern techniques to analyse stone tools and fossils. The 50th anniversary of Zinj's discovery is coming, and Domínguez-Rodrigo has a book being published.

John Kimaro, a Tanzanian antiquities officer overseeing Olduvai, says the government is considering guidelines for such situations, in which it seeks to maximize researcher access.

Meanwhile, observing researchers hope the years of research on 'the hallowed ground' of the Zinj fossils aren't harmed in the fray. ∎
Rex Dalton

S. WIDSTRAND/CORBIS

Biofuels receive boost from US government

Three new bioenergy research centres have each won $25 million per year for five years from the US Department of Energy to accelerate research into biofuels. Including external funding rustled up by the winning projects, the whole effort will top $500 million.

Three beaming investigators shook energy secretary Samuel Bodman's hand at a 26 June press conference in Washington DC. They represent three consortia, each of which includes at least one national lab. The lead institutions are Oak Ridge National Laboratory in Tennessee; the University of Wisconsin in Madison; and Lawrence Berkeley National Laboratory in Berkeley, California.

The new centres will have slightly different emphases, but all will take a crack at the famously difficult problem of coaxing ethanol out of cellulose — the tough indigestible parts of plants.

Asteroid strike may have formed Siberian lake

After nearly 100 years of searching, scientists have found what may be an impact crater made by the object that caused a huge blast over the remote Siberian area of Tunguska on 30 June 1908.

Most researchers think the blast, which flattened trees over an area of more than 2,000 square kilometres, was caused by an incoming stony asteroid exploding 5 to 10 kilometres above the ground. But numerous expeditions have failed to locate a crater or any fragments of the exploding object.

Now, Luca Gasperini from the Institute of Marine Science in Bologna, Italy, and others say that an almost circular lake called Lake Cheko may be a crater left by a remnant from that blast (L. Gasperini *et al. Terra Nova* doi:10.1111/j.1365-3121.2007.00742.x; 2007). Gasperini and his colleagues used seismic profiling, sonar and core sampling of bottom sediments to characterize the lake, which is about 8 kilometres from the epicentre of the explosion. Unlike other lakes in the region, 300-metre-wide Lake Cheko has a steep, funnel-shaped profile. The team plans to drill deeper into the lake bottom next summer.

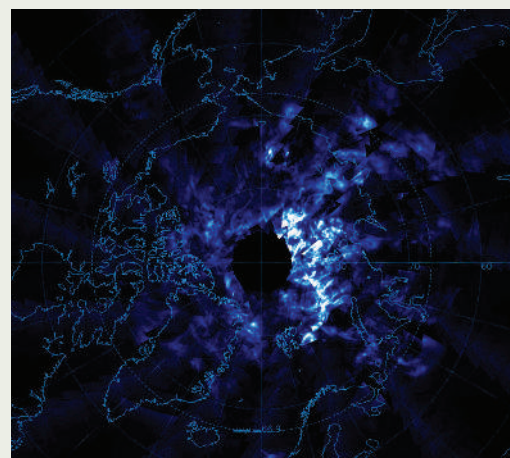## Ethical ruling and matured eggs offer hope for fertility

A Montreal mother has become the first woman to legally donate her frozen egg cells to her daughter. They are likely to be frozen for up to 20 years, until the daughter is ready to use them. She is now aged seven and has

AIM satellite snaps night-shining clouds

Noctilucent, or 'night-shining', clouds glow in shades of white and blue in this new image from the Aeronomy of Ice in the Mesosphere (AIM) satellite. Launched on 25 April, the mission snapped this picture on 11 June, as the season for seeing such clouds got underway in the Northern Hemisphere.

The clouds form in the mesosphere, about 85 kilometres up, and thus reflect sunlight long after the Sun has set — hence their name. They have been appearing more frequently and at lower latitudes in recent years, and some researchers suspect that may be due to climate change: as surface temperatures rise, those in the

mesosphere drop, enhancing the cold conditions needed for noctilucent clouds to form.

AIM, the first mission to study these clouds, will spend the next two years looking at them above both the Northern and Southern Hemispheres.

Turner syndrome, a congenital disease that causes ovaries to malfunction. The mother, in her early thirties, made "a very cogent argument" to the ethical review board of McGill University in Montreal, Canada, says Seang Lin Tan, the fertility expert who carried out the procedure.

Days earlier, Israeli doctors reported that they have successfully matured and frozen eggs from a girl aged five, potentially lowering the age at which cancer patients can preserve their fertility for future years. The work was led by fertility expert Ariel Revel of Hadassah University Hospital in Jerusalem.

Both results were presented at the European Society of Human Reproduction and Embryology's annual meeting in Lyon, France.

Health scare halts work on biodefence in Texas

The US government has suspended Texas A&M University in College Station from doing all federally funded biodefence research, after an oversight group found evidence that workers in its labs had been infected with biological-weapons agents.

On 26 June, the Sunshine Project based

Brucella bacteria have infected a Texas researcher.

in Austin, Texas, reported that three A&M researchers became infected with Q fever in 2006. Another lab worker was infected with the bacterium *Brucella* that same year, a fact not uncovered until April 2007. Both Q fever and *Brucella* are pathogens on the highly regulated 'select agents' list.
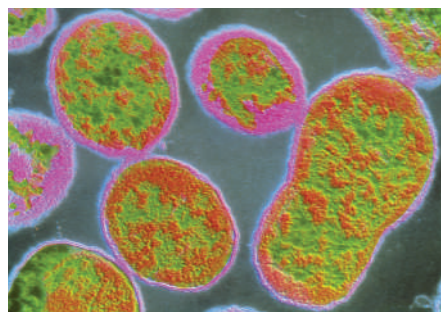
On 30 June, the US Centers for Disease Control and Prevention in Atlanta, Georgia, shut down all biodefence work at A&M, saying it will investigate the university's practices. A&M is currently on the shortlist to host a $450-million federal complex of biosecurity labs called the National Bio and Agro-Defense Facility (see *Nature* 442, 962–963; 2006).

US institute insists on open access to research

The Howard Hughes Medical Institute (HHMI), a leading private sponsor of biomedical research in the United States, will require its 300-plus investigators to make their research publicly accessible within six months of publication (see *Nature* 443, 894–895; 2006). Articles that do not meet this requirement will not be considered when the investigators apply for contract renewals.

The policy, announced on 26 June, will come into effect at the start of 2008 and will apply only to papers on which an HHMI investigator is the first, last or corresponding author.

HHMI investigators will still be able to publish in many publications, including *Science*, *Nature* and *Cell*. The institute has agreed to cover the costs incurred by some publishers, including Elsevier and John Wiley, for making the articles publicly available.

BUSINESS

Abbott's AIDS fight-back

Most drug companies have tried to avoid making enemies of AIDS activist groups. But Abbott Laboratories' patience has snapped, as **Erika Check** reports.

Early on in the war against AIDS, the drug industry learned that it was expedient to make nice with both developing-world governments and activist groups. But in an increasingly bitter dispute over access to one of its drugs in Thailand, Abbott Laboratories has taken a more aggressive approach.

In March, the Illinois drug firm withdrew its application to register the AIDS drug Aluvia in Thailand. And in May, it sued the French activist group Act Up-Paris in a Paris court after it had launched an attack on the company's website.

Abbott's hardball tactics — which have already attracted heavy flak — reflect the high stakes of the latest round of disputes to break out between the industry and governments over drug pricing and access. In the past, the access debate has focused mainly on access rights in very poor nations. "The drug-industry's strategy has been to ghettoize compulsory licences for poor countries and drugs for diseases that disproportionately affect the poor," says Robert Weissman, director of Essential Action, a non-profit group based in Washington DC that advocates wider access to patented medicines. "Now Thailand, which is not a lower-income country, is taking action on drugs for diseases such as heart disease and cancer, so the industry perceives that there is so much at stake."

Passing judgement

Since last November, the Thai government has issued a series of compulsory licences to enable it to manufacture or purchase generic copies of patented medicines from third parties without the consent of the patent holder. World Trade Organization agreements allow countries to issue such licences in "national emergencies" or "circumstances of extreme urgency". Needless to say, governments and industry disagree over exactly what these words mean.

In November, Thailand issued a compulsory licence for efavirenz, an AIDS drug sold by Merck. Two more licences were issued in January — one for Kaletra, the other for a heart drug marketed by Bristol-Myers Squibb and Sanofi-Aventis. Thailand is also said to be considering issuing compulsory licences for cancer drugs.

Both efavirenz and Aluvia are used in combi-

Up in arms: activists in Chicago join global protests over drug licensing.

nation with other AIDS medicines, and Aluvia — a heat-stable form of the drug Kaletra — could be a crucial weapon against AIDS in Thailand, where many citizens don't own refrigerators. But unlike Merck, which has been negotiating with the Thai government, Abbott reacted by indicating in March that it would not henceforth apply to sell new drugs in Thailand at all.

On 26 April — the eve of Abbott's annual shareholders' meeting — Act Up-Paris organized a 'netstrike' to protest against Abbott's announcement. Activists temporarily disabled Abbott's website by flooding it with requests. On 23 May, Abbott responded by suing Act Up-Paris under the French criminal code.

Abbott has justified its action by saying that it is ready to tolerate protests and demonstrations against its policies, but that the netstrike was illegal, and needs to be challenged. Drugmakers might also see the tactic as a good deal more menacing than, say, picketing of their shareholders' meetings, as it is has the potential to seriously disrupt their businesses.

Neither Abbott nor the drug-industry trade group the Pharmaceutical Research and Manufacturers of America (PhRMA) in Washington DC were willing to publicly discuss their strategies for dealing with the access issue, or with AIDS activist groups, in detail. But in a teleconference with reporters in May, PhRMA president Billy Tauzin told reporters that "it's

not just what Thailand is doing", pointing out that the industry is concerned about what it sees as a spreading epidemic of loss of respect for intellectual-property rights around the world.

And Abbott's actions in Thailand and France signal the depth of the company's concern over the licensing situation worldwide, says Ellen t'Hoen, from the Paris branch of Doctors without Borders (MSF). "These are acts of intimidation," she says, "and they are intended to make other countries think twice before they issue compulsory licences."

Indeed, the conflict has already spilled far beyond Thailand. In April, for instance, the office of the United States Trade Representative added Thailand to a Priority Watch List of companies viewed as lacking proper enforcement of international intellectual-property rules. But European countries and the Joint United Nations Programme on HIV/AIDS have supported the Thai stance. And on 20 June, Representative Henry Waxman (California, Democrat) and 34 of his colleagues sent the trade representative a letter protesting Thailand's placement on the watch list.

Setting a precedent

Even though everyone in the industry has strong interest in the outcome, other companies have been content to lay low while Abbott draws all the public attention. "While Abbott are taking these positions, the rest of them are having a quiet time," says Paul Cawthorne, head of mission for MSF Belgium's programmes in Thailand. "I think they're happy to let Abbott behave like a bull in a china shop."

But other companies are likely to be drawn into the fray as drug makers seek to defend their positions in places where immense poverty sits uneasily next to an increasingly prosperous middle class, which they view as a potentially lucrative new market. Novartis, for example, is now challenging India's new patent law in court, alleging that it doesn't go far enough to protect branded medicines. ∎

> "These acts are intended to make countries think twice before issuing compulsory licences."

See me here, see me there

Fifty years ago, a physics student dissatisfied with the standard view of quantum mechanics came up with a radical new interpretation. **Mark Buchanan** reports on the ensuing debate.

n 1957, a young physicist from Princeton University published his first paper — it went virtually unnoticed — and then disappeared from academia. He worked as an engineer and analyst in the defence industry until he died in 1982, at the age of 51. But Hugh Everett's lasting contribution to science, some physicists argue, stands far out of proportion to his paper tally. His first paper, they say, provided a new way to understand one of the most enduring puzzles of quantum physics.

Quantum theory has had many spectacular successes, but physicists have always been unsettled by its logical consistency. The theory, exemplified in Schrödinger's wave equation, insists that quantum particles such as electrons evolve into weird states of 'superposition' in which, among other possibilities, they can be in two places at once. This equation helped to explain the behaviour of atoms, but ordinary objects such as chairs are also made of quantum particles, so why don't we ever see them in two places at once?

One answer to this conundrum — as expressed in the late 1920s by Niels Bohr and Werner Heisenberg in their famous Copenhagen interpretation — was that we don't see these weird states because they collapse whenever we try to measure them. Everett, in bold contrast, suggested another solution — that the superpositions do affect our world, we simply don't notice them. As he pointed out, the maths of quantum theory suggests that when we encounter an object of superposition of say, here and there, that superposition draws us in too; splitting us into one being who sees the object here, and another who sees it there. In essence, as a later physicist put it, Everett claimed that quantum physics reveals a Universe that perpetually splits into "many worlds" co-existing side by side.

This idea was largely dismissed as being too weird, and many alternatives have been

KAPUSTA

suggested since. Empirical tests are unhelpful. Critics argue that experiments alone cannot distinguish between the many-worlds idea and some of the alternatives. "If experiments continue to verify quantum theory, we're going to be in a very difficult position, having to decide between theories not on evidence but on something else," says physicist and philosopher David Albert from the University of Columbia in New York. Supporters, though, argue that experiments are indeed helping their case — by continuing to find no evidence for the mysterious 'collapse' required by the Copenhagen interpretation.

Later this year, a select gathering of a few dozen physicists, philosophers and mathematicians will meet at two conferences in Canada and England to explore the status of Everett's ideas, 50 years on. "The idea," says philosopher of science Simon Saunders of the University of Oxford, and co-organizer of the meeting in England, "is to assemble the key arguments for and against the interpretation, and to come to a verdict — if not on the interpretation itself, then at least on the precise nature of its strengths and weaknesses."

Out of control

The interpretations of quantum theory — and the puzzle over why we never experience superpositions — has challenged physicists ever since the theory became indispensible for explaining atomic and nuclear physics early in the twentieth century. Bohr and the early interpreters of quantum theory got around this puzzle simply by dividing the world into classical and quantum parts, and supposing that quantum theory applied only to the latter. By their thinking, electrons and other quantum particles should evolve with quantum wave-like dynamics, but only until they interact with a classical object, such as a measuring device, at which point any superpositions would collapse, leaving just one outcome. Bohr pointed to the inherently uncontrollable nature of quantum interactions to account for the unpredictability of actual measurements.

In the 1930s, Hungarian mathematician John von Neumann formalized these ideas in his 'collapse postulate', which subsequently became one of the standard rules of quantum mechanics. Using the postulate, one can calculate the probabilities of the different outcomes when a 'measurement' takes place, with incredible accuracy. But quantum

theory still does not explain in precise detail when or why collapse should occur. For all its success, the theory seemed incomplete. "As they appear in the textbooks", says physicist Wojciech Zurek of the Los Alamos National Laboratory in New Mexico, "the axioms of quantum theory are inconsistent."

Bridging the divide

And for many physicists, most prominently Albert Einstein, Bohr's split between classical and quantum regimes introduced an unacceptable and arbitrary divide in physicists' view of the world. Quantum theory ought to be able to make sense of classical measuring devices, because these too are made of quantum particles. Restoring such wholeness has been the

Hugh Everett (right) challenged Niels Bohr's interpretation of quantum physics in his 1957 paper.

aim of quantum theorists ever since.

Starting in the 1950s, with American physicist David Bohm, theorists have tried their best to develop a coherent quantum theory. Bohm proposed a so-called hidden-variables interpretation, in which quantum particles have unique positions and velocities at all times. Bohm's theory doesn't require collapse, but its unusual mathematical structure — and the apparent impossibility of ever measuring these hidden variables experimentally — has made it attractive to only a few physicists.

Over the past two decades, various researchers have proposed interpretations that effectively side with Bohr's view, but make it more specific. These theories propose that Schrödinger's equation should be modified so that superpositions collapse naturally and, importantly, very rapidly in large, classical objects — thereby providing a more precise explanation of why we never see such superpositions at our level. They do away with the collapse postulate by effectively bringing it into the Schrödinger equation in a mathematically consistent way.
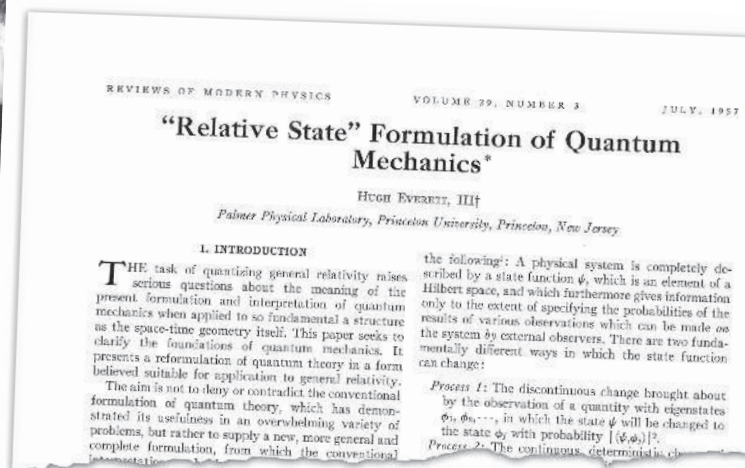
Everett's alternative takes yet another perspective. He simply insisted that collapse never happens, and that everything in the Universe always runs in tune with the wave-like dynamics of the Schrödinger equation. We seem to see things that look like collapse, he argued, only because we're part of the quantum world too, and so get caught up in its branching superpositions. Take quantum theory seriously, Everett insisted, and it supplies its own interpretation of reality — with the Universe splitting into multiple parallel worlds.

"In this view," says theorist David Deutsch from the University of Oxford, "our Universe is only a tiny facet of a larger multiverse, a highly structured object that contains many universes. Everything in our Universe — including you and me, every atom and every galaxy — has counterparts in these other universes."

All three approaches have their adherents, but for what seems to be a growing number of physicists, especially those working in quantum information and cosmology, it is Everett's alternative that wins out. "These fields follow the idea that there's nothing wrong with taking quantum theory to its logical limits," says Zurek. "And it was Everett who gave us the permission to do this."

Two-timing particles

Everett's idea is gaining popularity, at least in part, because quantum theorists have made progress in solving some of its earlier problems, which were not long ago thought to be fatal. One problem is that although Everett's view implies that the world splits into multiple branches, it does not, on its own, give a clear recipe for how this branching should take place. In standard quantum theory, for example, an electron in a fuzzy ill-defined

state can be thought of as a superposition of multiple states of well-defined position, or, alternatively, as a superposition of states with fuzzy positions but well-defined velocities. Neither picture is more correct than the other. Similarly, Everett's theory doesn't provide an obvious right way to identify the multiple co-existing branches that make up a superposition.

Deceptive behaviour

A body of ideas known as decoherence theory was developed by physicist Dieter Zeh of the University of Heidelberg in Germany in the 1970s, and has been expanded on by Zurek since then. The theory notes that because quantum systems interact with their environment they do not remain in superpositions for long, but instead tend to 'decohere'. In effect, environmental interactions make quantum systems behave as if their superpositions had collapsed, when in fact they have just become so entangled with the surrounding environment that no experiment, for practical purposes, would be able to detect them; and this is especially true of systems that involve many particles.

Many physicists suggest that, when coupled with decoherence, the many-worlds idea provides an attractive package. That's because decoherence singles out the familiar states of classical objects as being more robust and therefore observable. The idea even has some experimental support, and seems to remove much ambiguity from the many-worlds view.

Another long-standing problem for Everett's view is probability, one of the bedrocks of quantum physics. Using the collapse postulate, the magnitude of the wavefunction can be used to calculate probabilities for the position or velocity of an electron. But the many-worlds perspective, with its insistence on the never-collapsing evolution of quantum systems, says nothing at all about probabilities. "In Everett," says philosopher of physics Wayne Myrvold of the University of Western Ontario in Canada, "there seems to be no room for statements of probability at all."

Even so, recent work by Deutsch and by philosopher of physics David Wallace, also at the University of Oxford, suggests that quantum probabilities can find a way into the theory — through consideration of how human brains, or those of other organisms, register changes in their environments.

"It was Everett who gave us the permission to take quantum theory to its logical limits."
— Wojciech Zurek

They used decision theory — the logical science of optimal decision making — to explore how an individual living in a branching many-worlds universe would behave if he or she were trying to anticipate real-world outcomes as accurately as possible. They conclude that such individuals, confronting new outcomes as they entered new branches, would be driven to assigning probabilities to those branches, and would end up using the very same probabilities prescribed in the collapse postulate.

Although not everyone is convinced, some physicists think that Deutsch and Wallace succeeded in showing how the probabilities of quantum theory arise naturally from Everett's theory. "I'm not fully convinced that it all makes sense," says Myrvold, "but great strides have been made, and it's less implausible than I used to think."

Bigger and better

Everett enthusiasts also point to the increasing sophistication of modern experiments, which continue to verify the predictions of quantum theory, with no known exceptions. Using the techniques of atom trapping and quantum optics, experimentalists have observed quantum superpositions in systems containing more and more particles, as Everett's theory predicts ought to be possible. "The many-worlds view is slowly becoming the majority preference among physicists," says Tony Sudbery, a mathematician at the University of York, UK, "at least in part because we see quantum theory working, without any hint of collapse, for systems that are larger and larger."

But Albert says that the idea that experiments lend support to the many-worlds view is wishful thinking. As he points out, all experiments that verify quantum mechanics are consistent with the many-worlds theory and with several of the alternatives, including Bohm's interpretation of hidden variables. Observing wavefunction collapse in an experiment would contradict the Everett view, but the failure to do so might simply reflect the dif-

"Our Universe is only a tiny facet of a larger multiverse."
— David Deutsch

ficulty of doing controlled experiments on quantum superpositions.

Physicists hope that future experiments will challenge the idea of wavefunction collapse more directly by testing some of the theories that modify the Schrödinger equation. "Unfortunately," Albert says, "at this point, we're really very far from being able to do the necessary experiments." The current record for a quantum superposition is around 1,000 particles, although some researchers have bold plans to search for superposition collapse in a macroscopic object, perhaps a virus containing as many as 10^6 particles.

Seeing is believing

This is where the debate at the two meetings this year is likely to kick off: with arguments over what experiments can ever prove and whether theorists have really fixed all the conceptual holes in the many-worlds perspective. Although both events have been organized by Everett enthusiasts, they're unlikely to be one-sided affairs, as many researchers still have grave doubts about his ideas. "Even if one accepts 'many worlds'," says physicist Roger Penrose of the University of Oxford, "one needs a theory of consciousness, in effect, to explain the physics that we actually perceive going on in the world." Without it, he argues, the many-worlds theory is putting the cart before the horse.

Even Albert, who is lukewarm about the many-worlds view but positive about recent theoretical work, says, "My own guess is that, in the end, it's not going to work. But at least the people nowadays visibly understand what the worries and problems are."

Saunders and other Everett supporters suggest that those who hold out against the many-worlds picture are often driven, whether they say so or not, by an instinctive dislike of its non-intuitive consequences. "Quantum theory under the Everett interpretation is fantastic — too fantastic for the great majority of physicists to take seriously," Saunders says. But however they do it, as physicists struggle with the many-worlds interpretation, and the many alternatives, they are at least paying respect to the central idea of Everett's thesis: to take quantum theory very seriously indeed. ∎

Mark Buchanan is a science writer based in Europe. His latest book is *The Social Atom*.

See Editorial, page 1, and Commentary, page 23.

The biologists strike back

Time machines, spaceships, atomic blasters — the icons of science fiction tend to come from the physical sciences. But science fiction has a biological side too, finding drama and pathos in everything from alien evolution to the paradoxes of consciousness. *Nature* brought together four science-fiction writers with a background in the biological sciences to talk about life-science fiction.

Nature has given biologists a lot of weird things to study — how easy is it making aliens that are even weirder?

Peter Watts: Every now and then I think I'm taking a shot at it, but it always trips my ass ultimately. In my latest book, *Blindsight*, I thought I had come up with the ultimate alien: they didn't even have genes, and most of their metabolic processes were mediated by external magnetic fields. I thought I was really striking off into new and unexplored territory. But when it comes right down to it, I had described, at least in terms of gross morphology, brittle stars. Something that essentially had a whole series of pinhole cameras across its entire body surface, something like a very large telescope array.

Joan Slonczewski: I ended up in microbiology because microbes are the most fantastic creatures, more diverse than any other kind of creature in the natural biosphere. In fact, for much of my career it's been hard to tell which is more bizarre, the kind of research that I'm doing or the kind of science fiction I'm trying to write. In the textbook I'm doing now I've been writing about organisms that live 2 kilometres below the earth in gold-mines and that live off hydrogen atoms produced by uranium decay. I have yet to see nuclear-powered creatures much in science fiction.

Can science fiction work as a medium to put across important scientific ideas?

Joan: Yes. I actually teach a course on biology and science fiction for students who have trouble with a standard science course. There are certain books that do a wonderful job of teaching science through science fiction — Kurt Vonnegut's *Galapagos* is a great example.

Peter: I've got to second that. I think that was Vonnegut's best: it got evolution right. The idea that what is left of our civilization a million years hence is that when one of our seal-like descendents farts on the beach, the others just laugh and laugh — that's a wonderfully ironic and potent summation of human achievement.

Paul McAuley: Evolution is a keynote that runs through most of H. G. Wells's science fiction. The human race was going to slip down into unthinking Morlocks and Eloi or we could continue to rise and become the big-brained, small-limbed creatures that are the kind of epitome of science-fiction cliché of future man. Wells was taught by Huxley, had a zoology degree and so on, so he had a good grounding in it. But in Wells's time, evolution was some blind force. We've now got the opportunity to start directing evolution ourselves.

Joan: We can change our genes based on cultural views, what we believe are better genes. That's what the aliens in Octavia Butler's books are doing when they mate with humans. But what happens when it turns out the environment changes and that's not the best gene or we make a mistake? And what happens if we lose the variation that's required in the environment?

Those are great questions, but is science fiction good at answering them? Take cloning — how helpful has the vast amount of pre-existing science fiction about cloning been in informing the post-Dolly debate?

Joan: My impression is that for the students it was more helpful than not. That is, if Dolly had happened and there was no context at all, where would you begin to discuss what had happened? Whereas because there was a whole science-fiction tradition of questioning the ethics of cloning and the ethics of making people for spare parts, you had at least somewhere to start.

Ken Macleod: I think the prior art provided by science fiction was distinctly unhelpful in dealing with Dolly and cloning: ludicrous drek about cloning armies of soulless robots and *The Boys from Brazil* cloning Hitler, and the whole Frankenstein mess. And the actual ethical issues that arose with cloning were essentially none of the above.

Joan: I used to agree with your point that bad science fiction was an obstruction to learning, but as a biology professor I learned that sometimes bad science fiction is better than none at all. For example, you could poke Michael Crichton's portrayal of dinosaur cloning full of holes, but those stories encouraged a whole generation of molecular-biology students. As a result, we have a molecular-biology programme now at Kenyon College.

And that's really due to *Jurassic Park*?

Joan: Yes, we call it the *Jurassic Park* generation.

Paul: The big problem I have with Michael Crichton is he's basically anti-science. That old cliché of things that man wasn't meant to know embodies most of his work. Science is always running out of control, with people coming in to mop up afterwards. I think that the effect that Joan is seeing is from the film more than the novel, the wonderful scene of the dinosaurs up there on the screen. In the novel the hero was a lawyer. That's how anti-science it was.

My stuff gets compared to his stuff occasionally and I just have to say, 'Well, no'. Because I like science, I like scientists. I like what science does and I think that on the whole it's a good thing — and I think Crichton thinks the opposite, mostly. But the good thing about Crichton's work is that he does show, usefully I think, that science is not ethically valueless. Some scientists tend

Who's who

Ken Macleod took a master's in biomechanics and worked as a computer programmer before becoming a full-time writer based in Edinburgh. He is the author of the *Engines of Light* novels, the *Fall Revolution* novels and most recently of *The Execution Channel*.

Joan Slonczewski is a microbiologist and a professor of biology at Kenyon College in Gambier, Ohio. She is the author of six science-fiction novels, most recently *Brain Plague*.

Paul McAuley has been a full-time writer, almost exclusively of science fiction, since the late 1980s; before that he was a lecturer in botany at the University of St Andrews, Scotland. His latest novel is *Players*.

Peter Watts says he has a spent much of his adult life trying to decide whether to be a writer or a scientist, ending up as a marginal hybrid of both. Having done research in marine biology, he is the author of the *Rifters* trilogy and most recently of *Blindsight*.

to argue that knowledge is knowledge for knowledge's sake and that we should just find out what we can and damn the consequences. Crichton does actually cast the shadow of what we find back on to society and what's going to happen to it if we take these things to their logical conclusion.

Peter: When I start writing, I like to think of it as a sort of thought experiment. I go where the data lead, and I do not explicitly start off with a goal of writing a cautionary tale of saying that the world is turning to shit. That does seem to be where my stories all end up ultimately, but that's just because I'm following the data and there is an inertia to big systems and we can't realistically imagine a situation in which things would be better by 2050 unless we had actually had gotten serious about cleaning things up 20 years ago.

Ken: I think there's a distinction between science fiction and techno-thriller. The sort of thing that Michael Crichton writes is different from the sort of thing that Paul McAuley writes. Even when what Paul McAuley writes looks like a techno-thriller, he's actually sneakily writing science fiction in disguise. I attempt now and again to do the same thing. The difference is that in the techno-thriller, the lab eventually gets burned down, the genie gets back in the bottle, the evil scientist is defeated and so on. That's not the spirit of science fiction at all.

Paul: Ken's hit the nail on the head there. Science fiction posits that change is good and that change will happen and doesn't necessarily say in which direction change will go. The great power of science fiction is that, first of all, it's able to do that, and second, that it's able to get away without causing so much fuss. We can sneak in under the radar with all these outrageous notions and these manipulations and speculations about human nature.

Peter: Just to play devil's advocate for a moment, how would you guys react to the argument that if we can get away with it, we're not really doing the job? That if nobody gets pissed off by what you're doing, you are essentially conceding defeat when it comes to actually trying to provoke action.

Ken: I don't see science-fiction writers as agents for change. I think we're here not so much to bring about change but to investigate and imagine change. What happens as a result is the responsibility of the readers, not of the writers.
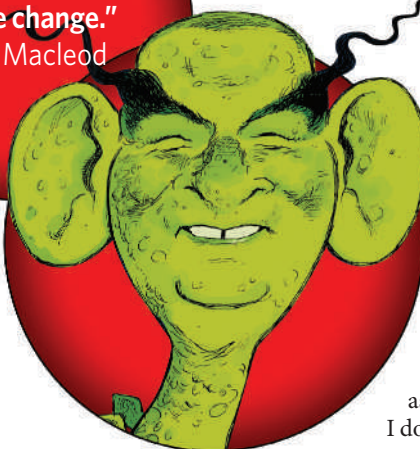
Joan: I think the imagination of change has to come first. If you don't do that, you're not a science-fiction writer. But I think there are some science-fiction writers who attempt to be scandalous. I think Robert Heinlein's *Stranger in a Strange Land* was considered extremely scandalous when it first came out — it posited a religion where there is cannibalism at the core of it, which was an attempt to imitate the Eucharist, and had all kinds of things that were then considered scandalous.

Ken: Yes, but have you ever come across anyone who was genuinely scandalized by *Stranger in a Stranger Land*? I mean, I was a little Christian fundamentalist when I first read it as far as I can remember and I wasn't scandalized. I was kind of mildly titillated, but…

Joan: In Ohio we're a little more easily scandalized, perhaps.

> "As a result of *Jurassic Park*, we have a molecular-biology programme at Kenyon College." — Joan Slonczewski

> "I think we're here to investigate and imagine change." — Ken Macleod

Science fiction has always been interested in 'the other' — and these days that other is as likely to be a computer program as an eight-legged alien. How does the interaction between biology, technology and artificial intelligence feature in your work?

Joan: One of the things that fascinates me is how people react to ideas of aliens or of artificial intelligence, and it seems to me that the way we treat artificial intelligences has a lot in common with the way we treat immigrant labour and the lower classes, or slaves. We think about these machines as slaves to do our work, but the more like us the machine is, the more effective a servant it is. This is the kind of dynamic we don't like to think about, but that has gone on in the way we treat either immigrant labour or slaves historically. Some studies have shown that even people who are very computer literate will treat their desktop computer as if it were another person. What if their computer were to become so powerful it actually wakes up and demands human rights?

In my book *Brain Plague*, there are technological entities that have sentience, whether they're robots or an entire transit system and it's just assumed that although they're robots, they're also just another ethnic group. My view as a molecular biologist is that our own bodies are machines composed of molecules and the computer on my desk is a machine composed of molecules, and the only difference is which one has woken up.

Ken: When I wrote my first novel, *Star Fraction*, I wrote it partly under the impulse of the feeling that Richard Dawkins and his selfish gene and the propagation of memes were something not widely enough known. I had to spread the word about these new and exciting ideas, and you know that whole thing of darwinian evolution going on in electronic systems seemed to me to be an enormously exciting and fruitful line of work and now it's pretty much all-pervasive on the talk shows. I think I was pretty much behind the curve even when I wrote it, though it didn't seem like it.

Peter: In my *Rifters* trilogy I wrote about Maelstrom, which was a far-future, massively super-evolved descendent of the Internet; all I did there was apply darwinian principles, assume you get your computer viruses, set them loose to breed in the wild, and end up with a seething electronic ecosystem that reproduces 200 times a second. It didn't strike me as a particularly radical innovation and I don't think any biologist would find that.

But people who were in AI [artificial intelligence] found this a massively innovative idea. I started getting letters from this guy who works in the Lawrence Livermore lab who told me that he had found my portrayal of digital ecosystems inspirational in his own work, which I found a little bit creepy because of the kind of things they do at the Lawrence Livermore labs and the fact that he couldn't tell me exactly what his work was.

The way my ideas about marine biology fall by the wayside and my ideas about AI get taken up makes me think that our imaginations are hamstrung in our own area of expertise. We know too many reasons why this, that or the other wouldn't work. We're perhaps a little too cognizant of our colleagues peering over our shoulders and ranking us as one or two steps above child pornographers because we write that sci-fi stuff in the first place. I wonder if some of the most innovative stuff comes when you retain the respect for logic and the respect for consequence, but you leave behind that infestation of fact and dogma that you used in getting your degree.

Paul: Well, that's why I quit science, folks. To get away from that self-censorship thing. I'm partly joking, but only partly. One of the useful things that science fiction does is to get out from under self-checking circuits that scientists must use when they're doing their work and just let rip and dance away with it. Doing science is like slogging through mud. Science fiction straps on mud shoes and dances off over the surface and onto the horizon, gesticulating madly and doing all sorts of silly little dances, but sometimes doing useful stuff.

Joan: I actually find science to be inspirational for science fiction. I can still remember

seeing an isolated photo-pigment that a grad student had got in a test-tube that was purple, and he shone light on it and it bleached white and this idea of the colourful switch enabled me to imagine: 'What if people had symbiotic microbes that would turn a switch depending on the environmental situation'. Later, that same pigment was used in a molecular switching device to make biochemical computers. So I think that science can be inspiring if you're doing it; you just have to be willing to not be inhibited in taking it a little farther.

Ken: In the novel I'm working on, one of the assumptions in it is that some AIs become self-aware because they're combat-robots and they're required to have ever more sophisticated theories of mind to work out what the guys they're about to shoot are going to do. But the other AIs, the ones that do our dirty work for us, like the police national artificial intelligence, which is one of the characters in my story, don't necessarily have self-awareness in the human sense at all.

Peter: The creepy thing about self-awareness is that 'the other' may in fact have been inside us all along, it may really be the one in control. The conscious decision to move your arm occurs half a second after the motor nerves have started firing, the conscious event is an executive summary received after the fact. This little self-aware homunculus behind the eyes doesn't seem responsible for nearly as much as it gives itself credit for; the heavy lifting seems to be done by something deeper, something we don't have conscious access to. I played with this idea in *Blindsight*, in which although there are aliens, there is no 'other'; the things our heroes meet are hyperintelligent but utterly nonsentient. And maybe that thing inside us that we can't feel, that makes the real decisions, that lets us think we're in control — maybe that's the same way.

> "The creepy thing abouty self-awareness is that 'the other' may have been inside us all along."
> — Peter Watts

> "Science fiction posits that change is good and that change will happen." — Paul McAuley

Let's end with your favourite biological moment in science fiction.
Paul: In *Blood Music*, by Greg Bear, when the protagonist looks down the microscope and he sees the bacteria have created little circuits that look like cities. A moment of 'Wow' like that is quite rare in science fiction, even though the 'Gosh, wow' thing is something we all aspire to. And it happened really early on in the novel, as well, which got even weirder after that. So I knew I was in for a good time.

Peter: Mine was Alice Sheldon's *The Screwfly Solution*. The idea is a rampant, literally epidemiological spread of homicidal hatred towards women, which society insists on treating as mass hysteria, whereas in fact it's been pheromonely introduced by aliens who want to clean up the real estate without using radioactive devices. So it's essentially a form of biological pest control.

Joan: For me, if it's a defining moment, it's the moment in Vonnegut's *Galapagos* in which the narrator of the story has the opportunity to decide whether to stick around for the next million years of evolution or to be taken off to heaven. And he decides that observing the next million years, no matter what, no matter how bad it is, that the next million years of human evolution are more compelling to him than going off to heaven. That to me is an inspiring moment.

Ken: I think that my sort of favourite biological science-fiction story is *Sunken Universe* [aka *Surface Tension*] by James Blish. It's an absurd idea that in the far future there are engineered human beings on another planet who are the size of protozoans. They're living in a puddle and they build what they call a spaceship, a little device made of bits of leaf and twig and so on that has wheels propelled by paramecium, and they laboriously drag this device across the dry land to the next puddle and at the end of it wonder if they have actually built this spaceship and crossed space like their ancestors did. And I loved that as an image of where we are and what we can do. ■

See Editorial, page 1. An expanded version of this conversation is at http://tinyurl.com/224s24

Not so good when 75% of grant applications fail

SIR — Your Editorial 'Never had it so good?' (*Nature* **447,** 231; 2007) claims that British science is in "rather good shape". Those in British universities who apply for research grants might not fully agree.

In the case of the UK Biotechnology and Biological Sciences Research Council (BBSRC), the new Labour government in 1997 made more cash available. As a result, the success rate for grant applications went up to about 40% in 1998–2000. Since then, however, there has been a steady decline to the present success rate of about 25%, leaving 10–15% of applications rated of international quality but unfunded. During the same period, the number of applications to the BBSRC has increased from about 1,200 a year in 1998 to about 1,900 a year in 2005. This may reflect a decline in other sources of funding from bodies such as the Medical Research Council, or a preference among applicants for funders who, unlike charities, include an amount for overheads.

This situation is not good news for British science, in that about 75% of applications to the BBSRC are now rejected, representing a huge waste of effort and ideas.

One obvious solution is to put more money into the system, to increase the success rate among applications ranked as internationally competitive. Another suggestion is to change the system for submitting and assessing applications, placing more of the onus on the universities, perhaps via a quota system for applications. A third is to weight the system more in favour of applicants' published track record and less in favour of the proposed science (with a special track for first-time applicants).

Above all, let's try to do something about this crazy situation in which so many grant applications fail, with the result that so much time is wasted for applicants, reviewers and administrators.

Philip Strange
School of Pharmacy, PO Box 228, University of Reading, Whiteknights, Reading RG6 6AJ, UK

Admission that intelligent design is a religious view

SIR — The case of Guillermo Gonzalez being denied tenure at Iowa State University should figure prominently in court next time there is an attempt to introduce intelligent design into the school science curriculum. According to your News story (*Nature* **447,** 364; 2007), Gonzalez, as a proponent of intelligent design, is appealing against the decision on the grounds that his application was rejected because of his religious beliefs, rather than his science. Is there any better evidence that intelligent design, by the admission of its own supporters (when convenient), belongs in classes teaching religion, not science?

H. A. Lessios
Smithsonian Tropical Research Institute, Box 0843-03092, Balboa, Panama

Terrorists are activists who renounce non-violence

SIR — As faculty members whose research was affected severely by a 2001 firebomb attack by the Earth Liberation Front (ELF), we object to the assertion in your Editorial 'Unwise branding' (*Nature* **447,** 353; 2007) that charging ELF arsonists with terrorism could amount to erecting an "unbreachable wall" to dialogue between them and scientists.

The ELF and its sister the Animal Liberation Front (ALF) use violence against civilian targets to further a political viewpoint: this is the definition of terrorism. Calling the ELF and ALF terrorist organizations is a simple statement of fact. There is no need to mince words in a vain effort to placate groups whose members, through a dangerous combination of wilful ignorance and willingness to enforce their world view 'by any means necessary', eschew reason in favour of senseless violence.

Some ELF and ALF apologists believe that 'property damage' (including destruction of research buildings at universities) does not qualify as terrorism. Perhaps the ELF statement from which we quote below will give those apologists a glimpse of the perspective shared by those of us whose names and addresses have been posted on ELF or ALF websites (which link to instructions on firebomb construction and deployment). After the firebombing of a US Forest Service laboratory in Pennsylvania in 2002, the ELF declared: "segments of this global revolutionary movement are no longer limiting their revolutionary potential by adhering to a flawed, inconsistent 'non-violent' ideology. While innocent life will never be harmed in any action we undertake, where it is necessary, we will no longer hesitate to pick up the gun to implement justice, and provide the needed protection for our planet that decades of legal battles, pleading, protest, and economic sabotage have failed so drastically to achieve."

Simply put, ELF and ALF members are anti-science, anti-intellectual, anti-human fundamentalists — certain that they have privileged access to some universal truth, deaf to alternative arguments, blind to evidence and determined to intimidate those who disagree with them. They are self-righteous in firebombing the very institutions (such as ours) that sponsor research and open discourse to understand and improve the state of the Earth for all its inhabitants. ELF and ALF terrorists have built the wall of naive, intolerant fundamentalism between themselves and us — only they can breach it. Rational people are, and always have been, waiting on the other side in the hope of receiving some form of communication other than a bomb or a bullet.

Sarah Reichard*, Thomas M. Hinckley*, H. D. Bradshaw, Jr†
*College of Forest Resources, University of Washington
†Department of Biology, University of Washington, Seattle, Washington 98195, USA

Activists: arson risks killing innocent people

SIR — Your Editorial 'Unwise branding' (*Nature* **447,** 353; 2007) is against equating animal-rights activism with terrorism. In it you state that "there is no such objective thing as a terrorist". This statement is yet another example of the moral blindness invading public discourse in the United Kingdom.

It is straightforward to define terrorism in an objective and legally egalitarian manner, for example by defining as a terrorist any person who uses violence to further his or her ideology, without taking into consideration the likelihood that innocent people may be injured, maimed or killed by such violent acts. Arson fuelled by ideology would certainly fit this definition of terrorism, and the animal-rights arsonists discussed in your Editorial were apparently not deterred by the possibility that people might be injured or killed in the fires they set.

Your second concern, regarding "who will be willing to publicly break bread with a terrorist, reformed or otherwise", is answered by current reality in Northern Ireland, South Africa and other places around the globe.

Mike Fainzilber
Biological Chemistry, Weizmann Institute, Rehovot, 76100 Israel

Activists: some walls are not meant to be breached

SIR — Your Editorial 'Unwise branding' (*Nature* **447,** 353; 2007) stated: "We should avoid building an unbreachable wall between criminal activists and their victims." Am I mistaken, then, in the purpose for incarceration of criminals after conviction?

Beverly E. Barton
Department of Surgery, University of Medicine and Dentistry of New Jersey, 185 South Orange Avenue, Newark, New Jersey 07103, USA

COMMENTARY

Many lives in many worlds

Accepting quantum physics to be universally true, argues **Max Tegmark**, means that you should also believe in parallel universes.

Almost all of my colleagues have an opinion about it, but almost none of them have read it. The first draft of Hugh Everett's PhD thesis, the shortened official version of which celebrates its 50th birthday this year, is buried in the out-of-print book *The Many-Worlds Interpretation of Quantum Mechanics*. I remember my excitement on finding it in a small Berkeley book store back in grad school, and still view it as one of the most brilliant texts I've ever read.

By the time Everett started his graduate work with John Archibald Wheeler at Princeton University in New Jersey quantum mechanics had chalked up stunning successes in explaining the atomic realm, yet debate raged on as to what its mathematical formalism really meant. I was fortunate to get to discuss quantum mechanics with Wheeler during my postdoctorate years in Princeton, but never had the chance to meet Everett.

Quantum mechanics specifies the state of the Universe not in classical terms, such as the positions and velocities of all particles, but in terms of a mathematical object called a wavefunction. According to the Schrödinger equation, this wavefunction evolves over time in a deterministic fashion that mathematicians term 'unitary'. Although quantum mechanics is often described as inherently random and uncertain, there is nothing random or uncertain about the way the wavefunction evolves.

The sticky part is how to connect this wavefunction with what we observe. Many legitimate wavefunctions correspond to counterintuitive situations, such as Schrödinger's cat being dead and alive at the same time in a 'superposition' of states. In the 1920s, physicists explained away this weirdness by postulating that the wavefunction 'collapsed' into some random but definite classical outcome whenever someone made an observation. This add-on had the virtue of explaining observations, but rendered the theory incomplete, because there was no mathematics specifying what constituted an observation — that is, when the wavefunction was supposed to collapse.

Everett's theory is simple to state but has complex consequences, including parallel universes. The theory can be summed up by saying that the Schrödinger equation applies

> **"If we dismiss theories because they seem weird, we risk missing true breakthroughs."**

Is it only in fiction that we can experience parallel lives? If atoms can be in two places at once, so can you.

MIRAGE ENTERPRISES/RGA

at all times; in other words, that the wavefunction of the Universe never collapses. That's it — no mention of parallel universes or splitting worlds, which are implications of the theory rather than postulates. His brilliant insight was that this collapse-free quantum theory is, in fact, consistent with observation. Although it predicts that a wavefunction describing one classical reality gradually evolves into a wavefunction describing a superposition of many such realities — the many worlds — observers subjectively experience this splitting merely as a slight randomness (see 'Not so random', overleaf), with probabilities consistent with those calculated using the wavefunction-collapse recipe.

Gaining acceptance

It is often said that important scientific discoveries go though three phases: first they are completely ignored, then they are violently attacked, and finally they are brushed aside as well known. Everett's discovery was no exception: it took more than a decade before it started getting noticed. But it was too late for Everett, who left academia disillusioned[1].

Everett's no-collapse idea is not yet at stage three, but after being widely dismissed as too crazy during the 1970s and 1980s, it has gradually gained more acceptance. At an informal poll taken at a conference on the foundations of quantum theory in 1999, physicists rated the idea more highly than the alternatives, although many more physicists were still 'undecided'[2]. I believe the upward trend is clear.

Why the change? I think there are several reasons. Predictions of other types of parallel universes from cosmological inflation and string theory have increased tolerance for weird-sounding ideas. New experiments have demonstrated quantum weirdness in ever larger systems. Finally, the discovery of a process known as decoherence has answered crucial questions that Everett's work had left dangling.

For example, if these parallel universes exist, why don't we perceive them? Quantum superpositions cannot be confined — as most quantum experiments are — to the microworld. Because you are made of atoms, then if atoms can be in two places at once in superposition, so can you.

The breakthrough came in 1970 with a seminal paper by H. Dieter Zeh, who showed

that the Schrödinger equation itself gives rise to a type of censorship. This effect became known as 'decoherence', and was worked out in great detail by Wojciech Zurek, Zeh and others over the following decades. Quantum superpositions were found to remain observable only as long as they were kept secret from the rest of the world. The quantum card in our example (see 'Not so random') is constantly bumping into air molecules, photons and so on, which thereby find out whether it has fallen to the left or to the right, destroying the coherence of the superposition and making it unobservable. Decoherence also explains why states resembling classical physics have special status: they are the most robust to decoherence.

Science or philosophy?

The main motivation for introducing the notion of random wavefunction collapse into quantum physics had been to explain why we perceive probabilities and not strange macroscopic superpositions. After Everett had shown that things would appear random anyway (see 'Not so random') and decoherence had been found to explain why we never perceive anything strange, much of this motivation was gone. Even though the wavefunction technically never collapses in the Everett view, it is generally agreed that decoherence produces an effect that looks like a collapse and smells like a collapse.

In my opinion, it is time to update the many quantum textbooks that introduce wavefunction collapse as a fundamental postulate of quantum mechanics. The idea of collapse still has utility as a calculational recipe, but students should be told that it is probably not a fundamental process violating the Schrödinger equation so as to avoid any subsequent confusion. If you are considering a quantum textbook that does not mention Everett and decoherence in the index, I recommend buying a more modern one.

After 50 years we can celebrate the fact that Everett's interpretation is still consistent with quantum observations, but we face another pressing question: is it science or mere philosophy? The key point is that parallel universes are not a theory in themselves, but a prediction of certain theories. For a theory to be falsifiable, we need not observe and test all its predictions — one will do.

Because Einstein's general theory of relativity has successfully predicted many things we can observe, we also take seriously its predictions for things we cannot, such as the internal structure of black holes. Analogously, successful predictions by unitary quantum mechanics have made scientists take more seriously its other predictions, including parallel universes.
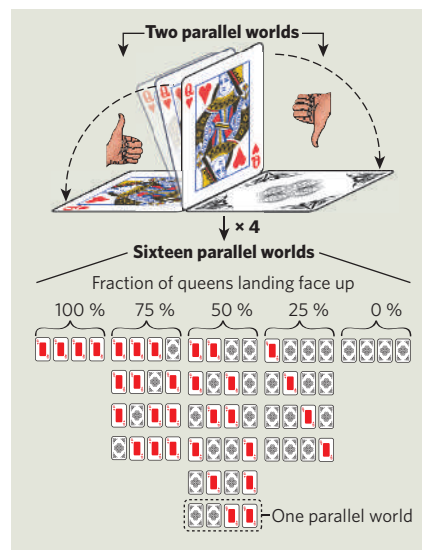
Moreover, Everett's theory is falsifiable by future lab experiments: no matter how large a system they probe, it says, they will not observe the wavefunction collapsing. Indeed, collapse-free superpositions have been demonstrated in systems with many atoms,

---

Not so random

According to quantum theory, a card perfectly balanced on its edge will fall down in what is known as a 'superposition' — the card really is in two places at once. If a gambler bets money on the queen landing face up, the gambler's own state changes to become a superposition of two possible outcomes — winning or losing the bet. In either of these parallel worlds, the gambler is unaware of the other outcome and feels as if the card fell randomly.

If the gambler repeats this game four times in a row, there will be 16 (2 x 2 x 2 x 2) possible outcomes, or parallel worlds. In most of these worlds, it will seem that queens occur randomly, with about 50% probability. Only in two worlds will all four cards land the same way up. If the game is continued many more times, almost every gambler in each of the worlds will conclude that the laws of probability apply even though the underlying physics is not random and, as Einstein put it, "God does not play dice". **M.T.**

Two parallel worlds

× 4

Sixteen parallel worlds

Fraction of queens landing face up

100 %   75 %   50 %   25 %   0 %

One parallel world

---

such as carbon-60 molecules. Several groups are now attempting to create quantum superpositions of objects involving 10^{17} atoms or more, tantalizingly close to our human macroscopic scale. There is also a global effort to build quantum computers which, if successful, will be able to factor numbers exponentially faster than classical computers, effectively performing parallel computations in Everett's parallel worlds.

The bird perspective

So Everett's theory is testable and so far agrees with observation. But should you really believe it? When thinking about the ultimate nature of reality, I find it useful to distinguish between two ways of viewing a physical theory: the outside view of a physicist studying its mathematical equations, like a bird surveying a landscape from high above, and the inside view of an observer living in the world described by the equations, like a frog being watched by the bird.

From the bird perspective, Everett's multiverse is simple. There is only one wavefunction, and it evolves smoothly and deterministically over time without any kind of splitting or parallelism. The abstract quantum world described by this evolving wavefunction contains within it a vast number of classical parallel storylines (worlds), continuously splitting and merging, as well as a number of quantum phenomena that lack a classical description. From their frog perspective, observers perceive only a tiny fraction of this full reality, and they perceive the splitting of classical storylines as quantum randomness.

What is more fundamental — the frog perspective or the bird perspective? In other words, what is more basic to you: human language or mathematical language? If you opt for the former, you would probably prefer a 'many words' interpretation of quantum mechanics, where mathematical simplicity is sacrificed to

collapse the wavefunction and eliminate parallel universes.

But if you prefer a simple and purely mathematical theory, then you — like me — are stuck with the many-worlds interpretation. If you struggle with this you are in good company: in general, it has proved extremely difficult to formulate a mathematical theory that predicts everything we can observe and nothing else — and not just for quantum physics.

Moreover, we should expect quantum mechanics to feel counterintuitive, because evolution endowed us with intuition only for those aspects of physics that had survival value for our distant ancestors, such as the trajectories of flying rocks.

The choice is yours. But I worry that if we dismiss theories such as Everett's because we can't observe everything or because they seem weird, we risk missing true breakthroughs, perpetuating our instinctive reluctance to expand our horizons. To modern ears the Shapley–Curtis debate of 1920 about whether there was really a multitude of galaxies (parallel universes by the standards of the time) sounds positively quaint.

Everett asked us to acknowledge that our physical world is grander than we had imagined, a humble suggestion that is probably easier to accept after the recent breakthroughs in cosmology than it was 50 years ago. I think Everett's only mistake was to be born ahead of his time. In another 50 years, I believe we will be more used to the weird ways of our cosmos, and even find its strangeness to be part of its charm. ■

In this Universe, Max Tegmark is a physicist at the Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

1. Shikhovtsev, E. *Biographical Sketch of Hugh Everett, III* http://space.mit.edu/home/tegmark/everett/ (2003).
2. Tegmark, M. & Wheeler, J. A. http://arxiv.org/pdf/quant-ph/0101077 (2001).

Surfing the multiverse

The 'many worlds' of quantum mechanics spawned many more of science fiction.
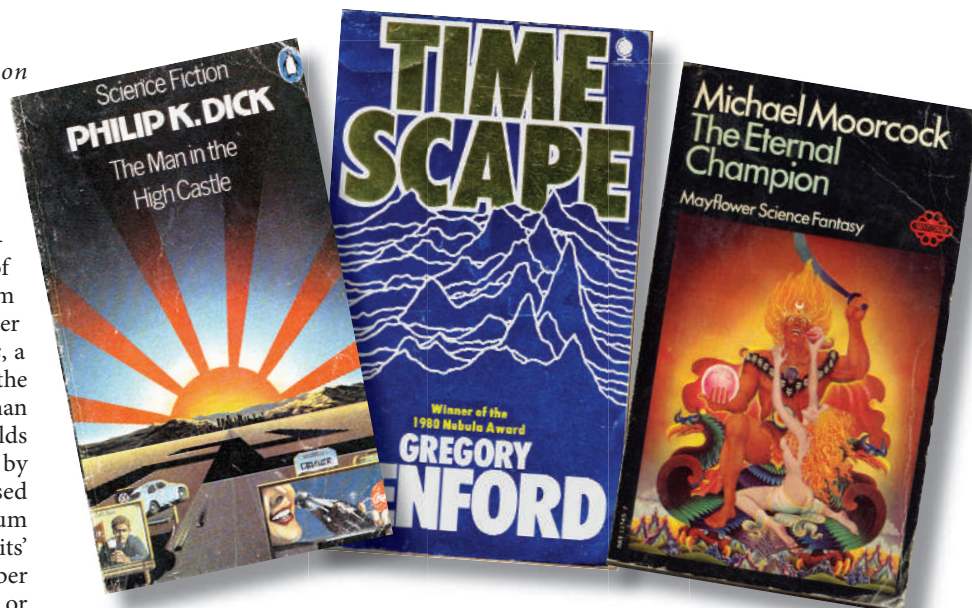
Gary Wolfe

"The futures we fail to encounter, upon the roads we do not take, are just as real as the landmarks upon those roads. We never see them, but we freely admit their existence…"

This passage reads today as a popularization of the many-worlds interpretation of quantum mechanics. In fact, it comes from 'Sidewise in Time', a tale by Murray Leinster published in 1934 in *Astounding Stories*, a leading pulp science-fiction magazine of the early twentieth century. This was more than two decades before the idea of many worlds was introduced into theoretical physics by Hugh Everett. His doctoral thesis proposed to resolve apparent paradoxes in quantum theory by suggesting that each event 'splits' the Universe, resulting in an endless number of separate, mutually exclusive histories or 'worlds' (see pages 15 and 23).

This notion has since become part of the fabric of science fiction and fantasy, ranging from absurdist jokes such as the "infinite improbability drive" in Douglas Adams's *The Hitchhiker's Guide to the Galaxy* to a magical means of travelling between worlds in popular series such as Stephen King's *The Dark Tower* or Philip Pullman's *His Dark Materials*. How did this meme take hold in literature and how has its use in fiction evolved?

Leinster, and Jack Williamson a few years later in his 1938 novel *The Legion of Time,* offered little scientific rationale for what was essentially a hot story-generating idea (although Williamson did presciently mention "an infinite proliferation of possible branches, at the whim of subatomic indeterminism"). But they were among the first to suggest that the idea might be somehow embedded in actual physics, rather than in such earlier occult notions as the 'astral plane'. For them, Everett's hypothesis must have seemed like a kind of validation. This was the tone of one of the first popular accounts of Everett's theory, a 1976 article in *Analog* — the same magazine, retitled, that had published Williamson and Leinster — by science-fiction writers Michael Talbot and Lloyd Biggle.

Writers outside science fiction viewed the possible multiplicity of worlds as more philosophy than physics, and were playing around with the idea well in advance of Everett. The most famous example, published in 1941, is probably Jorge Luis Borges's 'The Garden of Forking

Imagined worlds: which is stranger, science or fiction?

Paths' (in the collection *Labyrinths*), in which a famous maze-maker is described thus: "He believed in an infinite series of times, in a growing, dizzying net of divergent, convergent and parallel times… We do not exist in the majority of these times; in some you exist, and not I; in others I, and not you; in others, both of us."

Borges, like the science-fiction writers before him, was describing multiple timelines as an idea, as a potential condition of the Universe, not merely as a 'what-if' game or a story-making device. This "multiplicity of histories" — the term comes from science-fiction writer Stephen Baxter — is the essential nature of the true parallel-worlds tale.

More widespread are alternative histories. These have been traced back as far as the Roman historian Livy, who imagined what might have happened had Alexander the Great confronted the Roman Empire. Alternative or 'counterfactual' history has grown into such a bustling subgenre that it now has its own annual Sidewise Awards (in honor of Leinster). It virtually supports the careers of novelists such as Harry Turtledove, Eric Flint, Robert Conroy, and even Newt Gingrich and William Forstchen — all of whom have depicted alternative versions of the American Civil War or the Second World War. It also occasionally shows up on the bestseller lists, as with Philip Roth's *The Plot Against America*. This novel seems to be an autobiographical narrative about Roth's

own childhood in New Jersey until we realize that this is a world in which Charles Lindbergh defeated Roosevelt for the US presidency in 1940, and in which anti-Semitism is rapidly becoming state policy. Like other alternative histories, it posits only one divergence point and one alternative.

Possibly more fertile is a virtually infinite series of parallel worlds, occasionally impinging on our own. This has for decades provided the underlying conceit for English writer Michael Moorcock's vast body of work (ranging from the sword and sorcery of his 'Eternal Champion' tales to contemporary England in *The Cornelius Chronicles*). The Eternal Champion is a reluctant hero who finds his prowess tested in a variety of worlds. This is one of the most influential concepts in modern fantastic literature, echoed in such bestsellers as Robert Jordan's *Wheel of Time* series.

Moorcock had been developing his notion since his teenage years and was probably aware of Everett's thesis. "I would have come across a reference to it in *New Scientist* (one of my best friends was then editor) [or] physicist friends would have been talking about it," he has written. In 1962, Moorcock first introduced into science fiction the term 'multiverse', which was later used by physicist David Deutsch (in *The Fabric of Reality*) specifically in relation to Everett's thesis. A nice illustration of the symbiosis of science and science fiction.

Jack Williamson, *The Legion of Time* (1938). A pulp classic. The hero meets women from two possible futures that he may help determine.

Philip K. Dick, *The Man in the High Castle* (1962). Among the best and most influential alternative-world scenarios, in which the Allies lost the Second World War.

Michael Moorcock, *The Eternal Champion* (series, 1962–present). Several series of interconnected novels set in a multiverse of alternative realities.

Gregory Benford, *Timescape* (1981). One of science fiction's most compelling accounts of working scientists, dealing with cross-time communication to alter the past and thus create a variant future.

Stephen Baxter, *The Time Ships* (1995). This ambitious sequel to H. G. Wells's *The Time Machine* updates the classic tale with its notion of multiplicity of histories, and at one point directly alludes to Everett.

Among 'hard science-fiction' writers, quantum theory has become enormously influential. It provides something of a rationale for novels and stories that seek to question the fundamental nature of reality. Some of these writers go so far as to provide calculations and bibliographies to justify their fiction; others let their imaginations roam independently, more interested in metaphor than in theory.

For example, Frederik Pohl's *The Coming of the Quantum Cats* makes specific reference to quantum theory in a complex satirical tale. The main character encounters versions of himself from different time-streams — in one, Nancy Reagan is the US president, in another, Arabs rule the country. Meanwhile, the philosophical implications of quantum theory fascinate writers as much as physicists. Greg Egan's *Quarantine* begins as a detective thriller about a mental patient who can apparently walk through walls; it ends up as a challenging speculation on the role of the observer in the collapse of the wave function and how this might affect the many-worlds thesis. Towards the end of *Brasyl*, Ian McDonald suggests the mind itself may be a function of a kind of quantum computer entangled in a vast web of parallel universes — having introduced us to three disparate periods in Brazilian history, McDonald only gradually reveals the strange connections between them.

Such authors strive to incorporate actual developments in scientific theory into their fiction. They "play with the net up" in the words of science-fiction writer and physicist Gregory Benford. Among these authors, the many-worlds hypothesis has joined the vast arsenal of shaky but convenient speculations — along with time travel, faster-than-light propulsion, uploadable minds, quantum computing and alien contact. These speculations, not yet fully testable in reality, provide continuing fodder for the ongoing dialogue between scientifically literate fiction writers and practising theorists. Tracing the actual parameters of this dialogue over the past century of physics and fiction would be a daunting task, but it might well reveal, as Moorcock himself once put it, "the romantic imagination working … perfectly efficiently in both the arts and the sciences." ■

Gary Wolfe is professor of humanities and English at Roosevelt University, Chicago, Illinois 60605, USA.

EXHIBITION

Art shadowing science

Systema metropolis
by Mark Dion
Natural History Museum, London, UK
(15 June–2 September)

Colin Martin

American artist Mark Dion recently spent 18 months collaborating with scientists at London's Natural History Museum, collecting and identifying samples from four 'hidden' environments across the metropolis — cemeteries, a main road, a park and along the River Thames, including a power-station outlet.

"I'm not doing science," he says. "I shadow aspects of the process to get to an essence of the desire to do those things." The outcome of this investigation into the cross-fertilization between nature and urbanization in London is unveiled in an exhibition installed in five temporary pavilions in the museum's Jerwood Gallery. The scene is set with Dion's homage to Carl Linnaeus, whose binomial classification system democratized botany and zoology, enabling amateurs and professionals alike to collect and classify specimens. It includes a bust of the Swede and other historical material from the museum's own collection.

Photographs of the small invertebrates found at three cemeteries — including East Finchley where evolutionist T. H. Huxley is buried — are displayed in a simulated scientist's workplace (see picture). Another pavilion displays two square-metre samples of soil excavated from the 2012 Olympic Park. Fauna and flora colonizing

M. DION/NATURAL HISTORY MUSEUM

the soil will be kept alive throughout the exhibition, housed in a contemporary herbarium.

Insects were collected by a device mounted on a car while driving along one of London's busiest roads on 23 May, Linnaeus's birthday. The species of the squashed specimens were determined by DNA analysis and photographs of intact specimens of the species displayed.

Dion appropriates the methods, tactics and traits of science, to explore serious questions in a playful and subjective manner that scientists cannot adopt. In his own words, he is able to act as a "lightning rod" for ideas about science.

"I'm not conducting an experiment in strict scientific terms, but for some people it's enough to look the part," he comments in an interview published in the catalogue to a concurrent, touring exhibition, The Natural History of the Museum. This Dion retrospective charts his development as an artist, and is showing at the Dunkers Kulturhus in Helsingborg, Sweden, until 26 August (www.dunkerskulturhus.se) and at Seedamm Kulturzentrum in Pfäffikon, Switzerland, from 15 September to 11 November (www.seedamm-kultur.ch). ■
Colin Martin is a London-based writer.

Falling victim to balance

Balance: In Search of the Lost Sense
by Scot McCredie
Little, Brown: 2007. 304 pp. $24.95

Timothy Miles

Balance, or more precisely when people fall over because of imperfect balance, is an important issue for our ageing population. As the author of *Balance*, journalist Scott McCredie, points out, the social and medical costs resulting from falls are rising rapidly. One-third of people over the age of 65 and more than half of those aged 75 and over fall at least once a year, and many fall more frequently. A lot of these end up in hospital, and some never leave. Perhaps, as governments now begin to tackle the growing problem of obesity, they may also begin to think about strategies for tackling the increasing number of fall victims, which could be equally costly to society.

Scientists, pedantic folk that they are, will argue that the title of this book is misleading, given that 'balance' is not strictly a sense. Rather, balance is the outcome of an interaction of sensory signals that describe the orientation and movement of the head — which in turn stops one falling over. Thus, as the book is mainly about balance and not primarily about sensation, the reference to the "lost sense" seems rather inappropriate.

The emphasis is strongly on the workings of the vestibular system in the inner ear, with less discussion of the role of vision in balance and next to none on the important roles of sensory receptors in muscles and skin (collectively know as proprioception). This is a pity, as there is equally interesting information from both the laboratory and the clinic on the effects of disturbed proprioceptive inputs on balance and movement.

The book is wide-ranging in scope. There are many entertaining historical and anecdotal allusions ranging from Van Gogh's ear (sliced off in a futile response to intractable tinnitus or vertigo?) to The Flying Wallendas aerial balancing circus act. I was fascinated to read about the aptly but morbidly named 'lawn dart effect'. Apparently, when fighter pilots flying at low altitude accelerate horizontally at more than 1 g, the resulting 'off-the-scale' stimulation of the vestibular system can lead to the perception that the plane is climbing. In response, the pilot pushes the joystick forward to lower his trajectory and, well, yes, that's where the name came from.

The author is clearly intrigued by alternative (read unproven) therapies in which balance/vestibular-related interventions are used to treat a range of cognitive disorders including dyslexia and even autism. In one of these, hippotherapy, children with severe neurological disorders ride horses (rather than the large

Perfect poise: balance is as important for elderly people as it is for acrobats.

African mammals) to improve their "sensory integration". The author does qualify his description of these therapies by pointing out that most of them lack proper investigation and validation, although it remains to be seen whether this disclaimer would influence parents who are desperate for an effective therapy for their autistic child, for example.

Given the breadth of material covered, there are some surprising omissions of fascinating vestibular/balance-related phenomena. Among these is the ability to "steer" subjects by stimulating their vestibular system with weak electric currents. Turning on this stimulus in blindfolded people who are walking or pushing themselves in a wheelchair causes them to veer away from their target.

There are a number of errors in the science that will cause professionals in this field to mutter darkly. These are not hugely consequential, but it is a pity that they escaped pre-publication screening. However, my major concern with this book is the appendix, and in particular, the exercises that are suggested therein for improving balance. The book is obviously written for lay readers rather than health-care professionals, and it seems likely that it will appeal to patients with balance disorders or their carers who are seeking more information about a particular condition and on methods for addressing it.

The strong consensus among the therapists that I consulted is that elderly or balance-compromised patients should not attempt most of the exercises recommended in the appendix without professional supervision because of the high risk of falling while doing so: in the elderly, falling has a high probability of causing serious injuries such as a hip or wrist fracture.

I found *Balance* to be an entertaining read, and one that should appeal to the lay reader. However, it would benefit greatly from editing by a therapist, who would either axe the appendix or add a strong disclaimer, and a neuroscientist who would sharpen the science and encourage a more even-handed view of the roles of the many sensory receptors that contribute to balance. ■

Timothy Miles is a professor of physiology at The University of Adelaide's Medical School, Adelaide SA 5005, Australia.

H. TRYGG/GETTY IMAGES

Science with flare

The Sun Kings: The Unexpected Tragedy of Richard Carrington and the Tale of How Modern Astronomy Began
by Stuart Clark
Princeton University Press: 2007. 211 pp. $24.95/£15.95

Jan Stenflo

During the nineteenth century, astronomy was transformed from a mathematically oriented science preoccupied with mapping the positions and movements of celestial bodies to a physical science aiming to understand their nature and constitution. In *The Sun Kings,* science writer Stuart Clark offers a captivating account of this tortuous and passionately fought century-long transformation.

The tale centres around the solar astronomer Richard Carrington and his observations of the gigantic solar flare of 1 September 1859. Carrington, who today is little known outside the field of solar physics, happened to witness this event while drawing sunspots at his private observatory. Most solar flares can be seen only with modern spectroscopic equipment and so this white-light flare was very rare. The nearby Kew geomagnetic observatory recorded a weak magnetic disturbance lasting only 10 minutes, but exactly coincident with Carrington's flare. There was a violent geomagnetic storm 18 hours later and a worldwide display of the most magnificent aurorae, together with electric surges that disrupted telegraph communications.

Carrington's suggestion of a connection between the flare on the Sun and the geomagnetic disturbances and aurorae was largely

dismissed and ridiculed. Even 40 years later, Lord Kelvin, then president of the Royal Society, presented a 'proof' against a solar connection in an address published in *Nature*. Kelvin based his proof on Maxwell's electromagnetic theory to show mathematically how it would be energetically impossible for the Sun to cause the observed disturbances in the Earth's magnetic field. Around that time Kelvin also pronounced that "there is nothing new to be discovered in physics now. All that remains is more and more precise measurements."

The case for the solar connection was finally won in 1905, in a series of debates at the Royal Astronomical Society, thanks to the statistical analysis of Walter Maunder and the physical arguments of Joseph Larmor, the new Lucasian professor of mathematics at Cambridge. Larmor invoked the recently discovered electron to make the case that beams of charged particles from the Sun disturbed Earth's magnetic field. He suggested that "investigations of this rich tapestry of particle interactions beckoned, and would distinguish the twentieth century from the nineteenth".

One of the first to advocate turning astronomy into a physical science was the eminent astronomer William Herschel in the early

Carrington's private observatory in Redhill.

1800s. His call fell on deaf ears because it was related to another idea of his that was ridiculed, namely that the number of sunspots was statistically related to the price of wheat in England. Herschel was seeking a connection between sunspots and climate, which we now know exists. He used the price of wheat as a proxy for temperature, because there were no data on global temperature. Herschel found that when there were more sunspots, wheat was cheaper, implying warmer weather and bigger crops. Today's satellite observations show that the radiative output from the Sun is indeed higher when the sunspot number peaks.

The imprint of Carrington's 1859 flare could be seen in an ice core retrieved from Greenland

in 1992. The proton beams from solar super-flares produce nitrates in the Earth's atmosphere, which get deposited in the ice. Of all the flares of the past 500 years, the 1859 flare was one of the most powerful.

The most revolutionary tool for the physical exploration of stars was spectroscopy. Dark lines in the solar spectrum were noticed by William Wollaston in 1801 and rediscovered by Joseph von Fraunhofer in 1814, who later mapped these lines in much greater detail with his invention, the diffraction grating. Robert Bunsen and Gustav Kirchhoff showed that these absorption lines were the fingerprints of chemical elements — despite the famous French philosopher's Auguste Comte 1835 pronouncement that we would never know how to study the chemical composition of the stars.

Stuart Clark weaves all these events and ideas together in a fascinating tapestry. The account is accurate while being non-technical, and is suited for anyone with a general interest in the history of science. Clark's engaging writing style conveys the passion, intrigues and captivating life stories of the main players. It is a gripping tale of the birth of modern astronomy. ∎

Jan Stenflo is professor of astronomy at the Institute of Astronomy, 8092 Zurich, Switzerland.

A greener education

Degrees that Matter: Climate Change and the University
by Ann Rappaport and Sarah Hammond Creighton
MIT Press: 2007. 376 pp. $24.95, £15.95

Peter Hopkinson

Globally, higher education is a major service sector employing hundreds of thousands of people, educating millions of students, carrying out billions of pounds' worth of research and running huge campuses. These activities consume large amounts of fossil fuel and hence release significant amounts of carbon dioxide and other greenhouse gases. The central observation from *Degrees that Matter* is that universities are in a unique position to offer leadership on climate change and carbon emissions through their educational, research and wider roles in society.

This book is a case study of the attempts made at Tufts University, Massachusetts, to reduce carbon emissions through the Tufts Climate Initiative, a 15-year programme that began in 1991. The book was written by the initiative's directors and provides an exhaustive 330-page review and report on the actions taken to reduce carbon emissions at the university. Examples of initiatives include switch-off campaigns, procuring where possible renewable energy, promoting greener forms of travel and working with students and faculty mem-

bers to avoid energy waste. The good news is that the book has a wealth of ideas, recommendations and guidance.

For anyone who is responsible for energy or carbon management in a university, or is seeking to engage students and faculty in climate change, this book is compulsory reading. The chapters on personal action and climate change in the classroom are borne out of extensive experimentation and experience at Tufts and provide real stimulus for action within higher-education establishments.

The bad news is that despite the intense programme, carbon emissions at Tufts — both net and normalized — seem to have increased over time. The university as a whole has become more energy intensive, with the consequence that it will not meet its Kyoto target. This should, however, be set against other higher-education institutions, where the rate of increase over similar time periods is much greater and the reversal of trends, if at all, much slower. A large part of the increase is due to growing demands from personal equipment.

The picture may be even worse because the data do not take into account carbon emissions from staff, students and visitors commuting, or from business travel, which is widely accepted as having increased dramatically since 1990. The issue of travel is not given a great deal of attention in the book, although it is a source of considerable tension within any university

seeking to both promote 'internationalization' and reduce carbon emissions.

The one main weakness of the book is that there is relatively little empirical data on emissions sources, and little evaluation of the impact that the different initiatives at Tufts have had on emissions. The authors acknowledge the problems of measuring energy use and emissions, and argue that data collection, monitoring and reporting are crucial to carbon management and reduction. This is not an issue only at Tufts but is typical of most higher-eduction institutes where, historically, environmental-performance data have not been a priority, with a consequent lack of investment in systems for measuring and recording energy consumption .

With climate change and carbon emissions high on the political and educational agenda, the book is a timely reminder to the higher-education sector that it needs to attend to its own performance and reputation in this area. The evidence from Tufts shows the difficulties of reducing carbon emissions, but the Tufts Climate Initiative does provide a very good plan of action. The book's central message is that if universities wish to demonstrate practical leadership on reducing carbon emissions, there is a need for much greater action and progress than is currently found in most higher-education institutions. ∎

Peter Hopkinson is director of education for sustainable development at the University of Bradford, West Yorkshire BD7 1DP, UK, and co-director of Higher Education Environmental Performance Improvement (HEEPI).

ASTRONOMY

A constant surprise

Whether ancient or new, in distant galaxies or our own cosmic back-yard, stars have dramatic similarities that hint at remarkably robust formative processes.

John Cowan

The Big Bang gave us hydrogen, helium and a fraction of lithium. All the other elements in nature — the iron in our blood, the calcium in our bones and the gold in our jewelry — were synthesized inside stars that lived and died millions or even billions of years ago. In other words, the nature and extent of the synthesis of elements over the history of the Universe has changed with time and stellar evolution. Yet one aspect of stars has altered very little over billions of years, and of light years: the relative abundance patterns of certain heavy elements. These are consistent all the way through, from the Methuselahs of the star world to infants such as our Sun.

This suggests that the conditions forming these elements within stars have remained unchanged since the early Universe. Otherwise, today's stars would be producing a cocktail of elements different from the ancient stellar recipe. These striking similarities could help us to explore the nature of element formation throughout the Universe.

These findings answer a fundamental question of today's large surveys and detailed observations of the oldest surviving stars in our Galaxy and beyond. That is, whether the processes and stars that formed elements in our Galaxy were unique, or part of a broader pattern spanning large distances and times. Is our 'local neighbourhood' special, or really rather similar to other parts of the Universe?

Over the past decade, astronomical observations have unveiled interesting differences, but also pointed to commonalities throughout the Universe. For example, the oldest stars in our Galaxy lie in the galactic halo, the spherical 'cloud' of thinly scattered globular clusters and old stars surrounding it. What we are seeing is that the abundance pattern of rare heavy elements in halo stars — such as barium, europium and platinum — mirror those in our Solar System.

This is quite surprising, given the billions of years that elapsed between the birth of halo stars and our Sun. It implies that the formative processes for these elements,

including the types of stars and internal conditions, are remarkably robust.

Isotopic studies of certain elements back up this finding. Although it is much more difficult to obtain measurements of individual isotopes of a given element than the abundance of the element itself, recent observations of halo stars indicate that the abundances of europium and barium isotopes have also been consistent with the values we see in our Solar System. And the evidence that this commonality may extend beyond our Galaxy is growing.

Similarities in stars hold for neighbouring galaxies and beyond.

Studies of dwarf spheroidal galaxies have given us some tantalizing findings. These dim elliptical galaxies have star formation rates and histories distinct from those of the Milky Way, for instance leaving them with different proportions of high- and low-mass stars. Iron-poor stars in four dwarf spheroidal galaxies were observed to have heavy-element abundances similar to those of the Milky Way's halo stars. This is particularly striking at the very lowest and presumably earliest metallicities (or proportions of elements other than hydrogen and helium) where the abundance ratios of certain elements have similar values, and seem to be dominated by one formation process. Elements heavier in mass than iron are synthesized inside stars via nuclear reactions that involve a combination of neutron-capture processes. These

processes only occur on two timescales: rapid (much faster than a second) or slow (taking hundreds to thousands of years). Metal-poor stars in both the Milky Way and dwarf spheroidal galaxies are dominated by abundance patterns typical of the rapid process, which is thought to take place in the rare massive stars that explode as supernovae.

Is there any evidence that this commonality in stellar abundance patterns holds beyond our galactic neighbours? Yes. About 25 elemental abundances have been detected in a galaxy almost 12 billion light years away, formed a few billion years after the Big Bang. The number of heavy elements seen in it is relatively modest, but the abundance pattern is again consistent with the kind of relative or scaled abundance distribution we see in the Solar System. This is startling, and a strong indication that the solar abundances are, in many ways, cosmic.

A number of puzzles remain. We know, for example, that there are a few stars in the Milky Way with extremely low iron and extremely low heavy-element abundances. This suggests that some of the very first stars in our Galaxy were incapable of producing any significant fraction of heavy elements. So there may have been several generations of stars that formed early in our Galaxy. If so, how did they differ from each other? And there are certainly differences in iron abundances and star formation rates from galaxy to galaxy.

The abundance patterns we do see suggest more commonalities than differences, with neutron-capture element formation occurring everywhere early in the history of the Universe in a roughly similar manner. A focus on the similarities could give us a more coherent picture of our very interesting Universe.

John Cowan is David Ross Boyd professor in the Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma 73019, USA.

FURTHER READING
Cowan, J. J. & Sneden, C. Nature **440,** 1151 (2006).
Prochaska, J. X., Howk, J. C. & Wolfe, A. Nature **423,** 57–59 (2003).
Shetrone, M. et al. Astron. J. **125,** 684–706 (2003).

CONCEPTS

NEWS & VIEWS

MARS

Ancient fingerprints in the clay

David C. Catling

The thermodynamics of ancient clays on Mars seems inconsistent with the idea that a thick atmosphere of carbon dioxide caused a warm, wet era in the planet's early history. What did cause it remains an enigma.

Volcanic rocks dominate the surface of Mars. In the cold, dry conditions that prevail there today, these rocks can persist with little chemical alteration. But the recent discovery by Mars orbiters and landers of sulphate and clay deposits on the planet indicates that Mars' ancient environment was different, and involved liquid water[1]. On page 60 of this issue[2], Chevrier *et al.* add a new twist to this idea, with calculations showing how ancient martian clays formed in aqueous environments, but with little carbon dioxide present. That runs contrary to a popular view of an early Mars where liquid water was sustained by the greenhouse effect of a thick, CO_2-rich atmosphere.

Evidence that Mars was wetter before about 3.7 billion years ago comes from various geomorphic features[1]. In particular, valleys with characteristic branching forms seem to have been eroded by water, and some evidence also argues for the influence of rainfall. The heavily degraded rims of ancient craters and crater infilling are similarly interpreted as fluvial features in some models.

But whether a warmer, wetter earlier climate on Mars was persistent, intermittent or even existed at all remains controversial. If it did exist, what caused it? The idea that Mars' early atmosphere contained much more CO_2 than it does now is an appealing suggestion. More CO_2 provides greater warmth through its own greenhouse effect, together with that of increased water vapour. Because the early Sun was about 25% fainter around 3.5 billion years ago than it is today, several bars of CO_2 pressure would be required to achieve the necessary warming (1 bar is Earth's approximate atmospheric pressure).

At around 1.5 bar, however, CO_2 ice clouds start to form under the faint Sun. It has been suggested that these clouds might themselves produce warming, but up-to-date models show that CO_2 particles precipitate out or dissipate and cannot sustain a sufficient greenhouse effect[3]. Moreover, a substantial amount of atmospheric CO_2 should have left behind carbonate deposits. But on Mars' surface, not a single carbonate outcrop has been identified down to a horizontal scale of 100 metres. One possibility is that carbonates did not form

Figure 1 | Subsurface smectites. The yellow, reddish-brown and black veinlets in this view (about 1 mm across) of the 'Lafayette' martian meteorite are mixtures of clay and iron oxides that formed when small amounts of saline water infiltrated bedrock beneath Mars' surface and reacted with basaltic mineral grains. The presence of such clays beneath the martian surface is one caveat to interpreting Chevrier and colleagues' findings[2] as clinching evidence against a carbon-dioxide-rich early martian atmosphere. (Picture courtesy of A. Treiman, Lunar and Planetary Institute.)

because waters were rich in sulphuric acid, which displaces CO_2 (ref. 4). Although this is true in equilibrium, it still leaves the temporal evolution of CO_2 in the martian atmosphere unresolved.

Chevrier *et al.*[2] contribute to this debate by considering the thermodynamic equilibrium of clay minerals known as smectites. These clays, which have a characteristic layered-sheet structure, are found on Mars in both iron-rich and aluminium-rich forms. During the process of aqueous weathering, soluble ions are leached from 'primary' igneous rocks. 'Secondary' salts and clays, such as carbonates, sulphates and smectites, subsequently precipitate from the leaching fluid. What secondary minerals are deposited depends sensitively on the composition, pH and temperature of the leaching fluid. These minerals are consequently 'Rosetta stones' for deciphering past environments.

Chevrier *et al.* show that the equilibrium

between carbonates and smectites implies the presence of less than 0.001 to 0.01 bar CO_2 in Mars' early atmosphere for smectites to predominate as observed. Moreover, the neutral to alkaline conditions required to form the clays are inconsistent with the presence of sulphuric acid postulated to explain the absence of large carbonate deposits. According to this model, Mars' extensive sulphate deposits date from a later, more acidic geochemical era.

The lack of CO_2 in the early martian atmosphere could be explained by its early atmospheric loss. Mars' gravity is two-fifths that of Earth, making two processes, hydrodynamic escape and impact erosion, effective. Hydrodynamic escape would have occurred during Mars' first few hundred million years, when atmospheric hydrogen flowed out in a 'planetary wind' — analogous to the solar wind — that dragged along and removed heavier gases. Impact erosion results from the cumulative

effect of large impacting bodies releasing enough energy to blast parts of the atmosphere away. There is evidence for both processes in noble-gas isotopic ratios on Mars.

In the absence of CO_2, Chevrier et al. suggest[2] that other greenhouse gases promoted warmth. One candidate is methane; but methane is decomposed by ultraviolet sunlight, so a methane source comparable to Earth's biosphere would be needed to warm Mars above freezing. Another possibility is sulphur dioxide (SO_2), which, judging by the bulk chemistry of martian meteorites, could have been released from martian volcanoes in amounts similar to or exceeding their water emission. But SO_2 is soluble, and could have acted only as a 'lever' to raise temperatures to near freezing, making it easier for perturbations, such as asteroid or comet impacts, to cause temporary wet climates. Such impacts would have flash-heated the surface and released water, producing rain and erosion. Irrespective of any long-term greenhouse effect, the conclusion seems unavoidable that Mars was warmed transiently by many impacts early in its history[5].

A pivotal assumption made by Chevrier et al. is that the clays were formed in equilibrium with the atmosphere. But if the clays were formed in isolation from the atmosphere, beneath the surface, their thermodynamics might be of little relevance. Smectites are found in martian meteorites (Fig. 1) that formed in the subsurface when small amounts of saline waters infiltrated basalt[6]. Generally, suitable subsurface environments include long-lived hydrothermal systems that result from impacts on ice-rich ground; indeed, on Earth, smectites are characteristic of some hydrothermal systems in impact craters[7].

A further unresolved problem is that sulphates are found on Mars in places that range from young deposits around the planet's northern cap[8] to the ancient Meridiani outcrops[9], which are thought to be more than 3.7 billion years old. The stratigraphy of sulphate-rich deposits in an enormous chasm, Juventae Chasma, suggests that some of those deposits are similarly ancient[10]. The interpretation of Chevrier and colleagues' findings could therefore be more complicated than a geochemical history of an early age of clays succeeded by a sulphate era.

The history of recent Mars exploration shows that observations on the ground can completely overturn ideas inferred from orbit. Gusev Crater was chosen as a landing site for NASA's Spirit rover because it was interpreted from above as a lakebed. But what the rover found was a largely basaltic surface, rather than fluvial sediments. Meridiani Planum was selected for NASA's Opportunity rover because orbital data showed abundant haematite. But surface minerals there proved to be even richer in sedimentary sulphates.

Only a complete picture on the ground can provide confidence about the early environment of Mars, and whether it was ever conducive to life. Two rovers, NASA's Mars Science Laboratory rover and the European Space Agency's ExoMars, planned for launch in 2009 and 2013, respectively, might supply the answers — if targeted to the clays and sulphates of the planet. ■

David C. Catling is in the Department of Earth Sciences, University of Bristol, Queen's Road, Bristol BS8 1RJ, UK.
e-mail: david.catling@bristol.ac.uk

1. Carr, M. H. *The Surface of Mars* (Cambridge Univ. Press, 2006).
2. Chevrier, V., Poulet, F. & Bibring, J.-P. *Nature* **448,** 60–63 (2007).
3. Colaprete, A. & Toon, O. B. *J. Geophys. Res. Planets* **108,** doi:10.1029/2002JE001967 (2003).
4. Fairen, A. G., Fernandez-Remolar, D., Dohm, J. M., Baker, V. R. & Amils, R. *Nature* **431,** 423–426 (2004).
5. Segura, T. L., Toon, O. B., Colaprete, A. & Zahnle, K. *Science* **298,** 1977–1980 (2002).
6. Treiman, A. H. & Lindstrom, D. J. *J. Geophys. Res. Planets* **102,** 9153–9163 (1997).
7. Naumov, M. V. *Geofluids* **5,** 165–184 (2005).
8. Langevin, Y., Poulet, F., Bibring, J.-P. & Gondet, B. *Science* **307,** 1584–1586 (2005).
9. Squyres, S. W. *et al. Science* **313,** 1403–1407 (2006).
10. Catling, D. C. *et al. Icarus* **181,** 26–51 (2006).

SYNTHETIC BIOLOGY

Designs for life

Philip Ball

The genome of one bacterium has been successfully replaced with that of a different bacterium, transforming one species into another. This development is a harbinger of whole-genome engineering for practical ends.

If your computer doesn't do the things you want, give it a new operating system. As they describe in *Science*[1], Carole Lartigue and colleagues at the J. Craig Venter Institute in Rockville, Maryland, have now demonstrated that the same idea will work for living cells*.
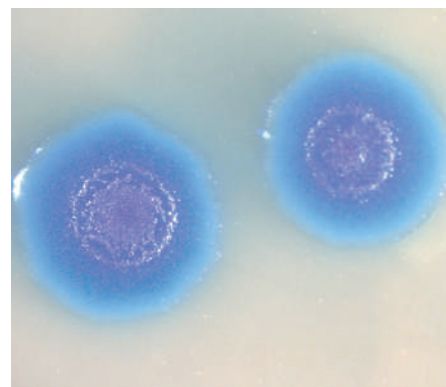
In an innovation that presages the dawn of organisms redesigned from scratch, the authors report the transplantation of an entire genome between species. They have moved the genome from one bacterium, *Mycoplasma mycoides*, to another, *Mycoplasma capricolum*, and have shown that the recipient cells can be 'booted up' with the new genome — in effect, a transplant that converts one species into another.

This is likely to be a curtain-raiser for the replacement of an organism's genome with a wholly synthetic one, made by DNA-synthesis technology. The team at the Venter Institute hopes to identify the 'minimal' *Mycoplasma* genome: the smallest subset of genes that will sustain a viable organism[2]. The group currently has a patent application for a minimal bacterial genome of 381 genes identified in *Mycoplasma genitalium*, the remainder of the organism's 485 protein-coding genes having been culled as non-essential.

This stripped-down genome would provide a 'chassis' on which organisms with new functions might be designed by combining it with genes from other organisms — for example, those encoding cellulase and hydrogenase enzymes, for making cells that respectively break down plant matter and generate hydrogen.

Mycoplasma genitalium is a candidate platform for this kind of designer-genome synthetic biology because of its exceptionally small genome[2]. But it has drawbacks, particularly a relatively slow growth rate and a requirement for complex growth media: it is a parasite of

*This article was published online on 1 July 2007.

Figure 1 | Genome swap. Colonies of the transformed bacteria (about 1 mm across).

the primate genital tract, and is not naturally 'competent' on its own. Moreover, its genetic proof-reading mechanisms are sloppy, giving it a rapid rate of mutation and evolution. The goat pathogens *M. mycoides* and *M. capricolum* are somewhat faster-growing, dividing in less than two hours.

Incorporation of foreign DNA into cells happens naturally, for example when viruses transfer DNA between bacteria. And in biotechnology, artificial plasmids (circular strands of DNA) a few kilobases big are routinely transferred into microorganisms using techniques such as electroporation to get them across cell walls. In these cases, the plasmids and host-cell chromosomes coexist and replicate independently. It has remained unclear to what extent transfected DNA can cause a genuine phenotypic change in the host cells — that is, a full transformation in a species' characteristics. Two years ago, Itaya et al.[3] transferred almost an entire genome of the photosynthetic bacterium *Synechocystis* PCC6803 into the bacterium *Bacillus subtilis*.

But most of the added genes were silent and the cells remained phenotypically unaltered.

Genome transplantation in *Mycoplasma* is relatively easy because these organisms lack a bacterial cell wall, having only a lipid-bilayer membrane. Lartigue *et al.*[1] extracted the genome of *M. mycoides* by suspending the bacterial cells in agarose gel before breaking them open, then digesting the proteinaceous material with proteinase enzymes. This process leaves circular chromosomes, virtually devoid of protein and protected from shear stress by the agarose encasement. This genetic material was transferred to *M. capricolum* cells in the presence of polyethylene glycol, a compound known to cause fusion of eukaryotic cells (those with genomes contained in a separate organelle, the nucleus). Lartigue *et al.* speculate that some *M. capricolum* cells may have fused around the naked *M. mycoides* genomes.

The researchers did not need to remove the recipient's DNA before adding that of the donor; instead, they added an antibiotic-resistance gene to the *M. mycoides* donor genome. With two genomes already present, no replication was needed before the recipient cells could divide: one daughter cell had the DNA of *M. capricolum*, the other that of *M. mycoides*. But in the presence of the antibiotic, only the latter survived. Some *M. capricolum* colonies did develop in the transplanted cells after about ten days, perhaps because their genomes recombined with the antibiotic-resistant *M. mycoides*. But most of the cells, and all of those that formed in the first few days, seemed to be both genotypically and phenotypically *M. mycoides* (Fig. 1), as assessed by means of specific antibodies and proteomic analysis.

The main question raised by this achievement is how much difference a transplant will tolerate. That is, how much reprogramming is possible? The DNA sequences of *M. mycoides* and *M. capricolum* are only about 76% the same, and so it was by no means obvious that the molecular machinery of one would be able to operate on the genome of the other.

Yet synthetic biology seems likely to make possible many new cell functions, not by whole-genome transplants but by fusing existing ones. When John Glass, a member of the Venter Institute's team, presented the transplant results at a recent symposium on the merging of synthetic biology and nanotechnology[4], he also described the institute's work on genome fusion (further comments on matters arising from the symposium appeared in last week's issue of *Nature*[5]).

One target is to develop a species of anaerobic *Clostridium* bacterium that will digest plant cellulose into ethanol, thus generating a fuel from biomass. Cellulose is difficult to break down — which is why trees remain standing for so long — but it can be done by *Clostridium cellulolyticum*. However, this creates glucose. *Clostridium acetobutylicum*, meanwhile, makes butanol and other alcohols, but not from cellulose. So a combination of

genes from both organisms might do the trick. For such applications, it remains to be seen whether custom-built vehicles or hybrids will win the race. ■

Philip Ball is a consultant editor for *Nature*.
e-mail: p.ball@nature.com

1. Lartigue, C. *et al. Science* doi:10.1126/1144622 (2007).
2. Fraser, C. M. *et al. Science* **270,** 397–403 (1995).
3. Itaya, M. *et al. Proc. Natl Acad. Sci. USA* **102,** 15971–15976 (2005).
4. Kavli Futures Symposium *The Merging of Bio and Nano: Towards Cyborg Cells* 11–15 June 2007, Ilulissat, Greenland.
5. *Nature* **447,** 1031–1032 (2007).

MOLECULAR MEDICINE
Entry granted

Edouard M. Cantin and John J. Rossi

The inability to efficiently deliver small interfering RNAs to target organs hinders their therapeutic application. So a demonstration of siRNA delivery to a notoriously difficult organ — the brain — is very exciting indeed.

The process of RNA interference (RNAi) uses small RNA triggers to decrease gene expression in a sequence-specific manner. In most applications, a small interfering RNA (siRNA) sequence — 21–23 nucleotides long — silences a target messenger RNA with a complementary sequence by either directing site-specific cleavage or inhibiting its translation into a protein[1,2]. This technique is a powerful and versatile tool for inhibiting gene expression, and has elicited a great deal of interest in harnessing it to treat human disease. Although during the past two years several significant findings have enabled systemic delivery of siRNAs to various tissues and organs[1], one tissue that has been difficult to target from the periphery is the brain; this is because traversing the blood–brain barrier poses a great challenge. On page 39 of this issue, Kumar *et al.*[3] describe a method to meet this challenge.

The junctions between the endothelial cells that line the brain capillaries prevent the passage of most molecules[4]. Chemically modified liposomes conjugated to monoclonal antibodies raised against epidermal growth factor can penetrate the mouse brain, as can

neuropeptides, but these approaches do not directly target neurons[4].

Similarly, viral vectors can be designed to express short hairpin-structured RNA sequences, which, once expressed, are processed to siRNAs. Such vectors have been used to introduce siRNAs directly into the brain tissue using the neurosurgical method of stereotaxis; thus, it has been shown that a decrease in the transcription of certain genes alleviates symptoms of neuronal diseases[5,6]. But direct injection within the cranium (the protective upper portion of the skull) is invasive, and systemic delivery of brain-targeting agents is a highly desirable alternative.

Kumar *et al.*[3] exploited the fact that neurotropic viruses — such as the rabies virus — that preferentially infect the nervous system can penetrate the brain. The rabies virus achieves this through glycoproteins on its lipid envelope. To test whether rabies viral glycoprotein (RVG) could promote transport of a non-rabies viral vector to the brain, the authors used the technique of pseudotyping, whereby an envelope glycoprotein from a specific virus is used to confer tropism for specific cell types to a viral
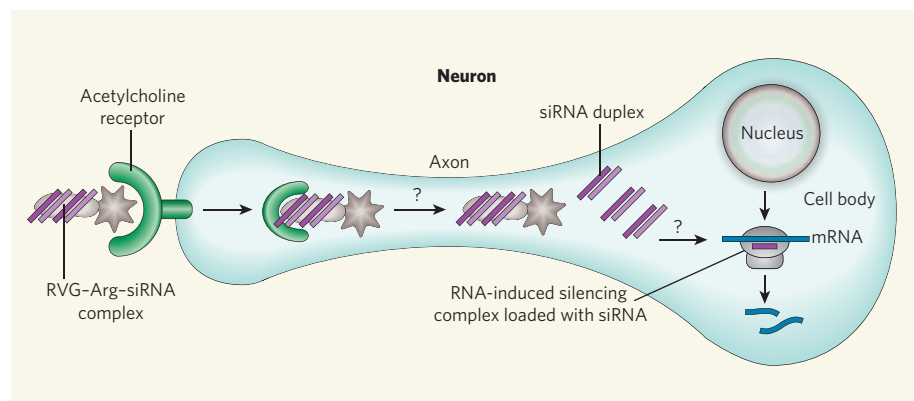

Figure 1 | Binding and uptake of peptide-conjugated siRNA by neuronal cells[3]. After binding to the acetylcholine receptor, the RVG–Arg–siRNA complex is internalized. It is then transported along the axonal process to the cell body, where siRNAs are released and incorporated into the RNA-induced silencing complex. Targeted cleavage of an mRNA sequence follows. The mechanisms of siRNA complex release from the acetylcholine receptor and its transport along the axon are not fully understood.

vector. They showed that RVG-expressing lenti-virus vectors could specifically infect neurons in culture, but not HeLa cells.

They then showed that a RVG-pseudotyped lentiviral vector, encoding a short hairpin RNA designed to target Japanese encephalitis virus (JEV), provided complete protection against this virus when both the vector and JEV were intracranially injected into the brains of mice. By contrast, the same lentiviral vector pseudotyped with the vesicular stomatitis G glycoprotein (VSV G) envelope did not afford protection, even though — as shown in previous work[7,8] — it reached neuronal cells. This is because retrograde spread of RVG occurs in axonal processes, facilitating its transynaptic spread to neighbouring neurons, whereas this doesn't happen for VSV G.

To transfer siRNAs into neural cells, Kumar *et al.* identified a 29-residue peptide from the RVG envelope that selectively binds to the acetylcholine receptor. They then fused this peptide with a sequence of nine arginine (Arg) residues that binds to negatively charged nucleic acids, including siRNAs[8]. They found that, when it was injected into mice, the peptide-conjugated siRNA (RVG–Arg–siRNA) specifically reached neurons (Fig. 1). This was detected through its effect on a synthetic protein known as enhanced green fluorescent protein (EGFP) that had previously been introduced into the mice.

The authors then questioned whether RVG–Arg–siRNA could be used to deliver siRNAs to brain cells *in vivo* after intravenous injection. They found that injection of mice expressing EGFP with RVG–Arg–siRNA complexes induced selective silencing of protein expression only in brain tissue. Injection of a control complex comprising the siRNA bound to the arginine sequence and the rabies virus matrix peptide (RV-MAT–Arg), however, did not affect EGFP expression. Similar results were obtained with an siRNA directed against the mRNA of the protein SOD1 — a natural target that is expressed in several cell types of the brain.

Finally, the authors used three intravenous injections of RVG–Arg–siRNA on successive days to protect mice from a JEV infection. Untreated mice, mice treated with the RV-MAT–Arg–siRNA complex, or those treated with a RVG–Arg–siRNA complex, in which the siRNA was a nonspecific sequence, all died within 10 days of JEV infection. However, 80% of mice treated with a siRNA complex specifically directed against JEV survived. By assaying for the levels of type I interferon, the authors ruled out the involvement of immune mediators, such as interferons, in protection against JEV infection conferred by RVG–Arg–siRNA complexes.

The widespread distribution of acetylcholine receptors in the thymus, the lungs and lymphocytes[9] doesn't seem to affect either the bioavailability or the specific delivery of the RVG peptide to the central nervous system.

But it is unclear which brain cells the RVG peptide reaches, because microglia, astrocytes, endothelial cells and neurons all express acetylcholine receptors[9]. Whether the RVG peptide traverses the blood–brain barrier to enter cells in the functional regions of the brain also remains unresolved. And which cells the siRNAs ultimately reside in is not known.

The exact mechanism by which the RVG peptide traverses the blood–brain barrier and enters cells is yet to be elucidated. However, the capacity of the gp120 protein of HIV to bind the acetylcholine receptor, and the ability of α-bungarotoxin and HIV Tat proteins to cross the barrier[8], suggest that peptides derived from other viruses that infect the brain — including HIV, the herpes simplex virus and flaviviruses — should also be explored as siRNA delivery agents.

Given the tremendous potential of RNAi as a therapy, it is tempting to speculate that the findings of Kumar and colleagues[3] could open the door for the treatment of neurological diseases. More immediately, however, it is hoped that the application of this approach to mice and large animal models of human diseases will demonstrate its usage, and indicate its feasibility, for human applications. ∎

Edouard M. Cantin and John J. Rossi are at the Beckman Research Institute of the City of Hope, and City of Hope Graduate School of Biological Sciences, Duarte, California 91010, USA.
e-mail: jrossi@bricoh.edu

1. Kim, D. H. & Rossi, J. J. *Nature Rev. Genet.* **8,** 173–184 (2007).
2. Martin, S. E. & Caplen, N. J. *Annu. Rev. Genomics Hum. Genet.* doi: 10.1146/annurev.genom.8.080706.092424 (2007).
3. Kumar, P. *et al. Nature* **448,** 39–43 (2007).
4. Pardridge, W. M. *Adv. Drug Deliv. Rev.* **59,** 141–152 (2007).
5. Davidson, B. L. & Boudreau, R. L. *Neuron* **53,** 781–788 (2007).
6. Xia, H. *et al. Nature Med.* **10,** 816–820 (2004).
7. Wong, L.-F. *et al. Hum. Gene Ther.* **17,** 1–9 (2006).
8. Wender, P. A. *et al. Proc. Natl Acad. Sci. USA* **97,** 13003–13008 (2000).
9. Gotti, C. & Clementi, F. *Prog. Neurobiol.* **74,** 363–396 (2004).

EVOLUTIONARY BIOLOGY

Mimicry on the edge

Thomas N. Sherratt

The latest turn in studies of mimicry in the animal world involves great tits as predators and almonds as prey. When it comes to being unpalatable, it seems that some mimics may neither flatter nor deceive.

Many species that are unpalatable to predators evolve conspicuous colour patterns to help advertise their unprofitability to would-be attackers. Other unpalatable species often mimic the same signals to share the costs involved in educating naive predators. Yet what selective forces operate when the defended mimics differ significantly in their unpalatability? This question has resulted in years of heated debate, but now Rowland and

colleagues (page 64 of this issue)[1] have tackled it experimentally.

The standard classification of protective mimicry, familiar to generations of biology students, was developed during the second part of the nineteenth century. Inspired by Amazonian butterflies, Henry Bates proposed that palatable species could gain protection from predators by resembling unpalatable species[2]. Bates also observed examples of unpalatable species

Figure 1 | Look-alikes. A vespid wasp (*Parachartergus* sp.; left) and a wasp-mimicking tiger moth (*Myrmecopsis strigosa*; right) from Guatemala. Many tiger moths are chemically defended and are therefore considered müllerian mimics of wasps[11]. Whether the better defended of the two species is helped or harmed by this mimicry is the general question investigated by Rowland *et al.*[1]. (Photo courtesy of R. Simmons and S. Weller.)

that closely resembled one another, but it was left to his contemporary Fritz Müller to provide a complete explanation. Müller's argument was based on 'strength in numbers' — unpalatable species may evolve similar appearances simply because this allows them to share the cost of teaching predators to avoid them[3].

Textbooks generally treat these two forms of mimicry as distinct phenomena, with deceitful batesian mimics exploiting the signals of unpalatable models, and honest müllerian mimics mutually reinforcing the meaning of their shared signals. However, generations of researchers have debated what type of mimicry occurs when the two or more unpalatable mimetic species do not have equal palatability[4,5], and whether batesian and müllerian mimicry represent endpoints of a continuum (Fig. 1). In particular, could a parasitic form of müllerian mimicry (quasi-batesian mimicry) arise, with weakly unpalatable mimics undermining the effectiveness of the signals of more unpalatable prey[4,6]?

Rowland et al.[1] now provide some empirical answers, but also highlight a complication long anticipated by theorists. In their experiments, the authors allowed individual wild-caught great tits (Parus major) to forage for prey in the form of pieces of almond. The pieces of nut were wrapped in paper marked with a simple symbol, and were distributed at random over an aviary floor. Sixty palatable prey were always provided. The rest of the prey consisted of systematic mixes of highly unpalatable 'models', moderately unpalatable mimics and entirely palatable mimics. High and moderate levels of unpalatability were generated by soaking the almonds in concentrated and dilute chloroquine phosphate solution, respectively. Like the models, the palatable mimics were always wrapped in paper marked with a solid black square, whereas the moderately unpalatable mimics carried either a solid black square (perfect mimics) or a curvier, diamond-shaped black symbol (imperfect mimics) (see Table 1 on page 64). In all instances, individual birds were allowed to forage for 50 prey items before a trial was complete.

Rowland et al. found that increasing the number of moderately unpalatable mimics reduced the overall proportion of highly unpalatable models that were attacked, whether the mimic was perfect or imperfect. Likewise, the presence of the highly unpalatable model reduced the 'mortality' of moderately unpalatable mimics. From these results, it might at first seem that all types of unpalatable prey are engaged in a mutually beneficial müllerian relationship. Yet when the number of palatable mimics was increased, the mortality of the highly unpalatable model stayed approximately constant — thus, even the batesian mimic was non-parasitic.

What drives this tendency towards mutualism? Adding more mimics to the system inevitably reduces the per capita mortality of the entire prey population because a fixed number of prey items were always attacked. The impact of these simple dilution effects has rarely been explored experimentally, but has long been predicted by theory[7]. For example, mathematical calculations predict that increasing the density of a palatable (or moderately unpalatable) mimic while keeping the model density constant (as did Rowland et al.[1]), can increase both the mimic and model survivorship simply because more food is available[8].

In contrast to the findings of Rowland et al., the only other experiment[9] to investigate quasi-batesian mimicry found that increasing the density of moderately unpalatable mimics actually increased the mortality of the more unpalatable model. Yet in this earlier experiment the overall prey densities were kept constant, so any increased tendency of predators to attack models would not have been masked by increases in food availability. Nevertheless, this cannot be the whole story because, by statistically controlling for the likely effects of dilution, Rowland et al. found that an increase in the density of moderately unpalatable prey had no significant effect on the probability that the birds would attack an individual model on encounter.

So, like well-behaved intermediates, the moderately unpalatable mimics neither reduced nor enhanced the attack rates of predators on the models following encounter, but increasing mimic availability did reduce the overall death toll. Perhaps the educational benefits of two unpalatable prey types sharing the same signal were roughly cancelled out by hungry or insensitive birds attempting to eat moderately unpalatable, yet nutritious, mimics. If so, then mimicry can indeed be considered as a continuum. Whether real mimics exhibit such moderate levels of unpalatability to predators is another question, but it seems reasonable to suppose that some do.

Certain unconventional theories of mimicry have been based directly on dilution phenomena[10], but to understand the evolution of mimetic signals I believe it is generally more informative to control for its effects. For example, one reason for the interest in quasi-batesian mimicry is that it can potentially explain the co-occurrence of several distinct forms of the same species of a moderately unpalatable mimic[6]. Rowland et al.[1] did not compare the survival of a moderately unpalatable mimic with a rare mutant form that resembles a different unpalatable model. However, the fact that predator behaviour towards models was largely independent of the density of unpalatable mimics indicates that no selection for multiple forms of the mimic would occur in this system.

Whatever the underlying mechanism, Rowland and colleagues' results suggest a process of 'advergent' evolution[5], in which the more weakly defended mimic evolves to resemble the better defended model, without markedly changing the way predators treat the model. I suspect that many müllerian mimicry systems

50 & 100 YEARS AGO

may be the result of such evolution — although it may often be driven by differences in the relative frequency of the mimetic species, rather than differences in unpalatability. ■

Thomas N. Sherratt is in the Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada.
e-mail: sherratt@ccs.carleton.ca

1. Rowland, H. M., Ihalainen, E., Lindström, L., Mappes, J. & Speed, M. P. *Nature* **448,** 64–67 (2007).
2. Bates, H. W. *Trans. Linn. Soc. Lond.* **23,** 495–566 (1862).
3. Müller, F. *Zool. Anz.* **1,** 54–55 (1878).
4. Speed, M. P. *Evol. Ecol.* **13,** 755–776 (1999).
5. Mallet, J. *Evol. Ecol.* **13,** 777–806 (1999).
6. Speed, M. P. & Turner, J. R. G. *Biol. J. Linn. Soc.* **67,** 281–312 (1999).
7. Holling, C. S. *Mem. Entomol. Soc. Can.* **45,** 1–60 (1965).
8. Sherratt, T. N. *Oikos* **103,** 93–100 (2003).
9. Speed, M. P., Alderson, N. J., Hardman, C. & Ruxton, G. D. *Proc. R. Soc. Lond. B* **267,** 725–731 (2000).
10. van Someren, V. G. L. & Jackson, T. H. E. *J. Lepid. Soc.* **13,** 121–150 (1959).
11. Simmons, R. B. & Weller, S. J. *Proc. R. Soc. Lond. B* **269,** 983–990 (2002).

IMMUNOLOGY

How a T cell sees sugar

D. Branch Moody

T cells of the immune system recognize lipids, as well as peptides, extending our ideas about such target antigens. A crystal structure of a T-cell receptor docked to antigen shows how a sugar controls lipid recognition.

Most cellular receptors coevolve with their target ligands through repeated interactions over millennia. The immune system has no such luxury. When new infectious agents such as HIV or avian influenza strike a population, immune receptors must reliably respond to foreign molecules — known as antigens — on their first encounter. For the immune cells known as T cells, this remarkable feat is accomplished by scrambling and recombining the peptide sequences that comprise the α- and β-chains of the T-cell receptor (TCR). These chains pair in millions of combinations, generating αβ-TCR heterodimers that can respond rapidly to almost any infecting microbe.

Immunologists used to believe that all T cells recognize peptide antigens that have become bound to proteins of the major histocompatibility complex (MHC I and II). But it is now known that T cells also respond to lipid antigens displayed by CD1 proteins[1–3]. On page 44 of this issue, Borg *et al.*[4] report the crystal structure of an αβ-TCR bound to a glycolipid antigen that is in a complex with a CD1 protein*. These images show at the atomic level how a TCR recognizes the glycolipid's sugar, and provide surprises that confound expert prediction.

The activation of T cells by CD1 and lipids has captured the imagination of immunologists (and physicians) for two main reasons. First, nearly all existing vaccines, adjuvants and immunodiagnostic tests involving T cells are designed to stimulate MHC function and track responses to peptide antigens. But if human T cells commonly recognize CD1 and lipids in disease states, then current immunomodulation strategies are missing many therapeutic targets.

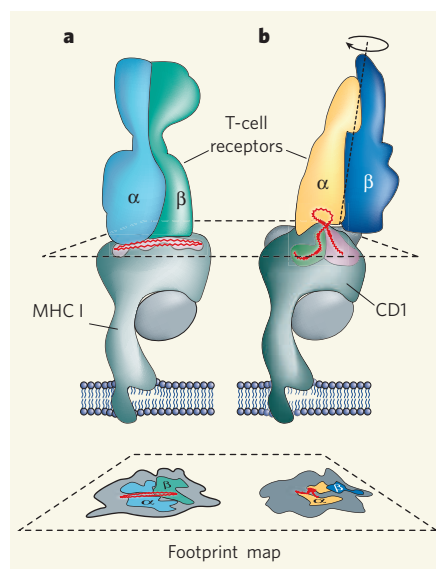The second, conceptual issue is that T cells

Figure 1 | Footprints of the T-cell receptor.
T-cell receptors (TCRs) are composed of α- and β-chains that bind to a composite surface formed by an antigenic molecule (red) in complex with antigen-presenting proteins from the MHC or CD1 families. **a**, In the classical mechanism of antigen recognition, both the α- and β-chains make contact with the antigen, so that their footprint covers the centre of the MHC platform. **b**, Borg and colleagues' crystal structure[4] shows that a CD1-reactive TCR is rotated clockwise from the classical position and pushed laterally so that only its α-chain makes contact with the antigen. This footprint shows how a TCR can recognize the carbohydrate structure in a glycolipid antigen.

bound to CD1, and so adapt the T-cell population after encountering lipid antigens. However, certain T cells, known as invariant NK T cells, are distinguished by their expression of nearly invariant TCR α-chains that recognize glycolipid antigens bound to CD1d (a member of the CD1 family). By sharing a limited receptor repertoire, many NK T cells can be activated in unison, eliciting particularly strong responses that have been shown to alter the outcomes of infectious, autoimmune, allergic, atherosclerotic and neoplastic[6] diseases in mouse models.

When lipid antigens for T cells were discovered, one commentator lamented that immunologists, who had only recently become comfortable with peptide chemistry, would next have to master lipids. But the situation was even worse than had been feared, as CD1-presented lipids are composed of lipid anchors conjugated to chemically diverse structures such as sugars, phosphates and peptides. The lipid chains of these antigens are inserted in a hydrophobic groove in CD1, allowing the more hydrophilic conjugated groups to protrude for recognition[7]. Glycolipids have emerged as major antigens, forcing biologists to consider how T cells recognize carbohydrates — structurally rigid rings of atoms connected by α- or β-linkages that define the orientations of the rings to each other.

The existence of glycolipid antigens makes immunological sense, because microbes synthesize lipids, sugars and peptides that differ markedly from those of vertebrates. Any of their characteristic chemical structures could be used by an organism to discriminate between its own and foreign matter. For example, NK T cells recognize CD1d bound to α-galactosyl ceramide (αGalCer), a glycolipid in which the sugar is in α-linkage with the lipid[8]. This antigen stimulates a potent immune response in mammals, but is considered 'artificial' because it is found in marine sponges that pose no threat to terrestrial vertebrates. However, certain bacteria also produce α-linked glycolipids[9,10], whereas the analogous mammalian glycolipids have sugars attached with β-linkages. Therefore, T cells might recognize α-linkages as markers of infection.

The structures of NK TCRs are known[11,12], so a sceptic might wonder what can be learnt from seeing how a known structure docks to CD1d. But Borg and colleagues' report[4] provides a wealth of fresh information. For example, the TCR forms contacts with CD1d at certain amino-acid residues that are shared with mouse CD1d, but are absent in human CD1a, CD1b and CD1c proteins. This explains the observation that NK T cells are specific for CD1d and why trans-species crossreactions have been observed.

Antigen motifs recognized by T cells have previously been described according to the position of a given amino-acid residue in the antigen's linear peptide chain. Borg *et al.* consider instead the point of attachment of groups to a sugar ring, and from this

*This article and the paper concerned[4] were published online on 20 June 2007.

responsive to CD1 proteins operate in a grey zone between the two major modes of immune recognition: innate and acquired immunity. CD1-reactive T cells might harness the power of rearranged TCRs — containing many different types of TCR α- and β-chain sequences[5] — to recognize almost any lipid molecule

PLANETARY SCIENCE

Hyperion the sponge

On 1 July 2004, seven years after its launch from Earth, NASA's Cassini orbiter fired its main engine, throttled back its speed and allowed itself to be pulled in by Saturn's gravity. In the three years since then, the spacecraft has provided snapshots of the strange and fascinating worlds that inhabit the Saturn system: the eerily Earth-like Titan, for instance, the rain on whose plains is mainly methane; or Enceladus, from near whose 'tiger-striped' southern pole an icy plume spouts into space.

No moon, however, seems quite as odd as the body that hove into Cassini's close view on 25 September 2005. Irregularly shaped and chaotically rotating, Hyperion, the eighth-largest saturnian moon, looks for all the world like a bathroom sponge.

Two papers in this issue take a detailed look at this odd character. In the first of these, Thomas *et al.* (P. C. Thomas *et al. Nature* **448,** 50–53; 2007) conclude that Hyperion's sponge-like appearance (Fig. 1 on page 51) comes from an unusually high density of well-preserved craters between 2 and 10 kilometres across. According to the authors' theory, these discrete craters

survived, rather than having been eroded away or filled in by material ejected in the impacts that formed them, because Hyperion's interior is indeed porous. A meteorite hitting a significantly porous body will compress its surface rather than excavate it, and produce much less ejected material.

The authors arrive at this explanation by first calculating Hyperion's mean radius — no mean feat, as its irregular shape and random spinning make imaging it in one shot impossible. They were also able to compute the moon's mass through a dynamical model requiring that Hyperion maintain a stable orbit in the complex gravitational environment of the Saturn system. This mass was only just over half that expected for a body of Hyperion's 135-km calculated mean radius, if it were made of water ice — and considerably less if higher-density materials are present. In other words, there is far less within Hyperion than had been expected.

In the second of the papers, Cruikshank *et al.* (D. P. Cruikshank *et al. Nature* **448,** 54–56; 2007) delve more deeply into Hyperion's make-up. Using data from ultraviolet and infrared spectrometers onboard

Cassini, they divide the moon's surface into distinct areas of high reflectivity, seemingly dominated by water ice, and low reflectivity, concentrated at the craters' bases. These latter areas have a considerably diminished water-ice signature, but a prominent absorption band best explained by the presence of solid carbon dioxide in complex with a further, unknown material.

Hyperion's spectra, like those of other saturnian satellites such as Phoebe and Iapetus, also provide evidence for irregularly scattered, nitrogen-rich, organic molecules. The image shown here of three

surveys overlaid on the moon's surface topography (blue, H_2O; red, CO_2; green, carbon nitrides, CN) gives an idea of the complexity of the picture. Cyan indicates areas with strong H_2O and CN signals, but little CO_2; yellow areas are dominated by CO_2 and CN, with little H_2O; and magenta equates to CO_2 and H_2O, but little organic material.

What these discoveries tell us of Hyperion's origin and the wider history of the Saturn system is as yet unclear. But as the pictures from Cassini continue to roll in, the varied worlds they depict do not cease to surprise.

Richard Webb

---

perspective speculate on what kinds of glycolipid other than αGalCer might promote the best TCR–CD1 interaction. Their structure also shows how the α-linked sugar lies down to allow a tight CD1–TCR interface, whereas an upright β-linked sugar might block the approach of the TCR to CD1d.

The TCR footprint on CD1 is perhaps the most pleasing aspect of this study, because it provides a simple explanation for the observation that nearly all NK T cells require the same kind of TCR α-chain to function[6]. Imagine that the α- and β-chains of TCRs are legs. MHC-reactive TCRs generally adopt a confident, two-footed stance, with both chains contacting their peptide antigen near the centre of the MHC platform. With some variation[13], this mode of binding holds true for 15 TCR–MHC structures[14] and was predicted to apply also to CD1. But the NK TCR is a ballerina in a pirouette: it leans sideways and has spun clockwise, contacting its antigen only with the α-chain, while using the β-chain to graze CD1 at its edge (Fig. 1). Whether other TCRs dance on one foot is not yet settled, as T cells responding to CD1a, CD1b and CD1c

do not seem to share the same type of α-chain for antigen recognition[5].

Another notable feature is the displaced location of the TCR β-chain — it is pushed to an extreme lateral position so that it hangs over the edge of CD1d. This lateral approach is reminiscent of a TCR found on γδ T cells[15]. Perhaps these two unusual footprints will serve as counterpoints to help explain how most MHC proteins guide TCRs to their centres[14].

During the long wait for the first TCR–antigen–CD1 crystal, specialists guessed that the TCR would align near the centre of CD1, as occurs with the MHC. This reasonable — but apparently mistaken — prediction did not arise from lack of consistency in MHC footprints, but rather from the faulty assumption that the behaviour of the MHC could be used to predict that of CD1. Such thinking has a chequered past, and dates back to the description of CD1 proteins as MHC-like molecules. This oft-used comparison provides a good introduction to CD1, but is superficial and even misleading. Modern evidence shows that these systems differ in their levels of polymorphism, mutability of antigens, loading mechanisms, trafficking

pathways and other features. CD1 is more than just the MHC for lipids, and its footprint provides another reminder that we should think about CD1 biology in its own terms. ■

D. Branch Moody is in the Department of Medicine, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, 1 Jimmy Fund Way, Boston, Massachusetts 02115, USA.
e-mail: bmoody@rics.bwh.harvard.edu

1. Calabi, F. & Milstein, C. *Nature* **323,** 540–543 (1986).
2. Porcelli, S., Morita, C. T. & Brenner, M. B. *Nature* **360,** 593–597 (1992).
3. Beckman, E. M. *et al. Nature* **372,** 691–694 (1994).
4. Borg, N. A. *et al. Nature* **448,** 44–49 (2007).
5. Grant, E. P. *et al. J. Exp. Med.* **189,** 195–205 (1999).
6. Cui, J. *et al. Science* **278,** 1623–1626 (1997).
7. Moody, D. B., Zajonc, D. M. & Wilson, I. A. *Nature Rev. Immunol.* **5,** 387–399 (2005).
8. Kawano, T. *et al. Science* **278,** 1626–1629 (1997).
9. Mattner, J. *et al. Nature* **434,** 525–529 (2005).
10. Kinjo, Y. *et al. Nature Immunol.* **7,** 978–986 (2006).
11. Gadola, S. D. *et al. J. Exp. Med.* **203,** 699–710 (2006).
12. Kjer-Nielsen, L. *et al. J. Exp. Med.* **203,** 661–673 (2006).
13. Hahn, M., Nicholson, M. J., Pyrdol, J. & Wucherpfennig, K. W. *Nature Immunol.* **6,** 490–496 (2005).
14. Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. *Annu. Rev. Immunol.* **24,** 419–466 (2006).
15. Adams, E. J., Chien, Y. H. & Garcia, K. C. *Science* **308,** 227–231 (2005).

ARTICLES

Transvascular delivery of small interfering RNA to the central nervous system

Priti Kumar[1], Haoquan Wu[1], Jodi L. McBride[2], Kyeong-Eun Jung[3], Moon Hee Kim[3], Beverly L. Davidson[2], Sang Kyung Lee[4], Premlata Shankar[1] & N. Manjunath[1]

A major impediment in the treatment of neurological diseases is the presence of the blood–brain barrier, which precludes the entry of therapeutic molecules from blood to brain. Here we show that a short peptide derived from rabies virus glycoprotein (RVG) enables the transvascular delivery of small interfering RNA (siRNA) to the brain. This 29-amino-acid peptide specifically binds to the acetylcholine receptor expressed by neuronal cells. To enable siRNA binding, a chimaeric peptide was synthesized by adding nonamer arginine residues at the carboxy terminus of RVG. This RVG-9R peptide was able to bind and transduce siRNA to neuronal cells *in vitro*, resulting in efficient gene silencing. After intravenous injection into mice, RVG-9R delivered siRNA to the neuronal cells, resulting in specific gene silencing within the brain. Furthermore, intravenous treatment with RVG-9R-bound antiviral siRNA afforded robust protection against fatal viral encephalitis in mice. Repeated administration of RVG-9R-bound siRNA did not induce inflammatory cytokines or anti-peptide antibodies. Thus, RVG-9R provides a safe and noninvasive approach for the delivery of siRNA and potentially other therapeutic molecules across the blood–brain barrier.

The endothelial cells of brain capillaries form extremely tight junctions, providing a superfine filter that prevents the transport of most molecules from the vasculature into the brain parenchyma[1–3]. To overcome this, the conventional approach in gene therapy experiments has been to inject siRNA and viral or DNA vectors directly into the brain by stereotactic surgery (reviewed in refs 2, 3). However, these methods result only in localized delivery around the injection site, with no widespread effects within the brain; they are also too invasive for human therapy. If one could overcome the blood–brain barrier (BBB), intravenous administration would provide the ideal noninvasive means for delivery throughout the brain because of the rich vascularity of the brain, with capillaries encasing virtually every brain cell[3].

Because neurotropic viruses do cross the BBB to infect brain cells, we asked whether the strategy used by viruses to enter the central nervous system could also be used to enable delivery of siRNA to the brain. We chose rabies virus to test this hypothesis because it shows a high degree of neurotropism *in vivo* and the cellular entry mechanisms have been well characterized.

RVG pseudotyping confers neuronal cell specificity

RVG interacts specifically with the nicotinic acetylcholine receptor (AchR) on neuronal cells to enable viral entry into neuronal cells[4,5]. We therefore initially tested whether pseudotyping lentivirus with RVG, instead of the conventionally used vesicular stomatitis virus glycoprotein (VSV-G), could confer specificity for neuronal cells. Green fluorescent protein (GFP)-encoding lentiviral vector Lentilox pLL3.7 (ref. 6) pseudotyped with either RVG or VSV-G was tested for its ability to infect neuronal or non-neuronal cells. Whereas VSV-G pseudotyped lentivirus infected both cell types, RVG pseudotyping resulted exclusively in the infection of Neuro 2a cells, not HeLa cells (Supplementary Fig. S1a). Because RVG has been shown to mediate retrograde axonal transport and increase the spread of a viral vector

within the brain[7], we also tested whether RVG pseudotyping of pLL3.7 encoding a short hairpin RNA (shFvE^J)[8] targeting Japanese encephalitis virus (JEV) increases its antiviral efficacy. Different concentrations of shFvE^J lentivirus, pseudotyped with RVG or VSV-G, were tested for protection efficacy in an intracranial JEV challenge assay[8]. Whereas at a high dose (2×10^5 transducing units) both lentiviruses afforded protection equally, at a lower dose (2×10^3 transducing units), all mice treated with RVG-pseudotyped lentivirus survived but all those treated with VSV-G-pseudotyped lentivirus succumbed to JEV infection (Supplementary Fig. S1b). Taken together, these results suggest that RVG confers neuronal cell specificity and in addition, by facilitating retro-axonal and trans-synaptic spread[7], enhances the transduction of neighbouring neuronal cells.

RVG peptide binds specifically to neuronal cells

The snake-venom toxin α-bungarotoxin (BTX) specifically binds to the AchR[9], and a short (29-residue) peptide derived from RVG competitively inhibits the binding of BTX to the AchR in solution[10]. We reasoned that this peptide might bind specifically to neuronal cells expressing the AchR. To detect binding, we synthesized a biotinylated RVG peptide or a control peptide of similar length derived from the rabies viral matrix protein (RV-MAT). When tested for cell binding, RVG peptide was found to bind to the AchR-expressing Neuro 2a cells[11,12] but not to the receptor-negative HeLa cells, whereas RV-MAT peptide bound to neither cell type (Fig. 1a). RVG peptide also did not bind several other non-neuronal cells tested (Fig. 1b). To confirm AchR-mediated binding specificity, we tested whether BTX could inhibit RVG peptide binding to Neuro 2a cells. Indeed, BTX inhibited RVG binding in a dose-dependent manner (Fig. 1c). Moreover, BTX was also able to displace prebound RVG from Neuro 2a cells (not shown). Next we tested whether RVG peptide could also specifically bind primary neuronal cells. Freshly isolated mouse brain cells but not splenocytes bound the RVG peptide, and neither cell

[1]The CBR Institute for Biomedical Research and Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Department of Internal Medicine, Roy J. and Lucille J. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. [3]Research Center, Samchully Pharm. Co. Ltd., Seoul 135-735, Korea. [4]Department of Bioengineering and Hanyang Fusion Materials Program, Hanyang University, Seoul 133-791, Korea.
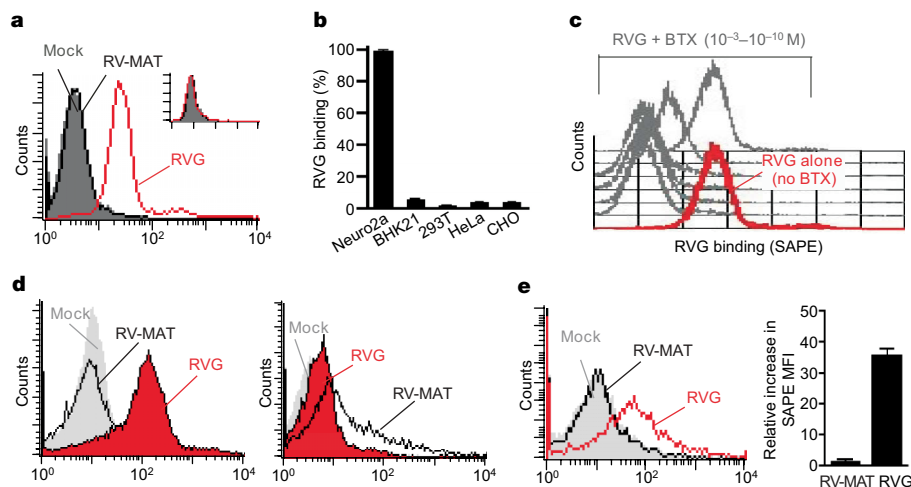
Figure 1 | A short RVG peptide binds to neuronal cells *in vitro* and *in vivo*. **a**, Neuro 2a and HeLa cells (inset) were incubated with biotinylated RVG or RV-MAT peptides, stained with SAPE and examined by flow cytometry. **b**, Peptide binding was also tested with the indicated cell lines in triplicate. Error bars indicate s.d. RV-MAT did not bind any of the cell lines (not shown). **c**, Neuro 2a cells were stained with biotinylated RVG in the absence (red histogram) or presence (grey histograms) of decreasing concentrations of BTX. **d**, Freshly isolated mouse brain (left) and spleen (right) cells were tested for peptide binding. **e**, Mice were injected intravenously with biotinylated RVG or RV-MAT peptide; 4 h later, isolated brain cells were stained with SAPE. Error bars indicate s.d. ($n = 6$).

type bound RV-MAT (Fig. 1d). Because AchR is widely expressed in the brain, including the endothelial cells of brain capillaries[13], we also examined the ability of RVG peptide injected intravenously to cross the BBB and enter brain cells. Mice were injected with biotinylated RVG or control RV-MAT peptides and, 4 h later, brain cell suspensions were examined by flow cytometry. As shown in Fig. 1e, brain cells from mice injected with RVG, but not those injected with RV-MAT peptide, were positive for peptide uptake, indicating that the RVG peptide might cross the BBB to enter brain cells.

RVG-9R peptide delivers siRNA to neuronal cells

Although RVG peptide can bind to neuronal cells, it does not bind nucleic acids and therefore cannot be used to transport siRNA. However, short, positively charged, cell-penetrating peptides bind negatively charged nucleic acids by charge interaction[14–16]. A nona(L-arginine) peptide was reported to be highly efficient in facilitating the cellular uptake of nucleic acids, and replacement of L-arginine with D-arginine (to form 9dR) enhanced the uptake even further[17]. Moreover, a cholesterol-conjugated oligo(D-arginine) has been used to deliver siRNA to a transplanted tumour in mice[18]. Thus, we tested whether RVG fused to 9dR could bind and deliver siRNAs to neuronal cells. For this we used RVG-spacer-9dR (designated RVG-9R) and control RV-MAT-spacer-9dR (RV-MAT-9R) chimaeric peptides. Both peptides were able to bind siRNA in a dose-dependent manner in a gel-shift assay (Fig. 2a). RVG-9R was also able to transduce fluorescein isothiocyanate (FITC)-labelled siRNA into neuronal cells in a dose-dependent manner and, in agreement with siRNA binding studies, a 1:10 molar ratio of siRNA to peptide was found optimal for maximal transduction (Fig. 2b). To determine the neuronal specificity of siRNA delivery, Neuro 2a and HeLa cells were transduced with FITC-siRNA complexed to RVG-9R or RV-MAT-9R, and Lipofectamine transfection was used as a positive control. Lipofectamine enabled siRNA uptake by both cells, and RV-MAT-9R was unable to transduce either cell type (Fig. 2c). In contrast, RVG-9R transduced Neuro 2a cells, but not HeLa cells, to a similar degree to Lipofectamine. Thus, RVG-9R allows neuronal cell-specific siRNA delivery.

Although RVG-9R could transduce siRNA to Neuro 2a cells in the above assay, siRNA is not functional unless it is delivered into the cytoplasm. Thus, we also assessed the gene-silencing ability of the siRNA delivered by RVG-9R. Neuro 2a cells stably expressing high levels of GFP were transduced with anti-GFP siRNA, bound to RVG-9R or RV-MAT-9R or transfected with siRNA by using

Lipofectamine, and GFP expression was determined 2 days later. RV-MAT-9R-complexed siRNA was unable to decrease GFP levels, whereas RVG-9R/siRNA silenced GFP expression to a similar extent to Lipofectamine transfection (Fig. 2d), suggesting that the RVG-9R-delivered siRNA was indeed functional. The RVG-9R/siRNA
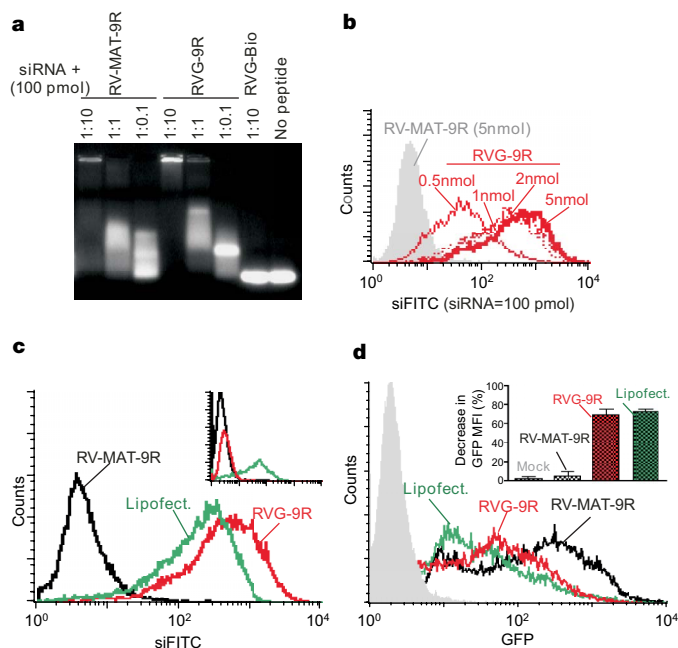

Figure 2 | RVG-9R peptide binds and delivers siRNA to neuronal cells *in vitro*, resulting in gene silencing. **a**, Mobility of free or peptide-complexed siRNA was analysed by agarose-gel electrophoresis. **b**, Neuro 2a cells were examined for uptake of FITC-siRNA complexed with RVG-9R at the indicated concentrations. **c**, Neuro 2a and HeLa (inset) cells were examined for uptake of FITC-siRNA complexed with RVG-9R or RV-MAT-9R peptides at a 1:10 molar ratio. Lipofectamine transfection (Lipofect.) was used as a positive control. **d**, Neuro 2a cells stably expressing GFP were transduced with GFP siRNA complexed with RVG-9R or RV-MAT-9R peptides, and GFP silencing was tested 2 days later. A representative histogram and cumulative data from three independent experiments (inset) are shown. The grey filled histogram represents Neuro 2a cells not expressing GFP. MFI, mean fluorescence intensity. Error bars indicate s.d.
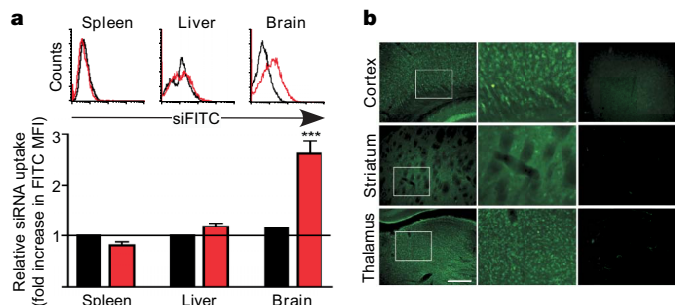
Figure 3 | RVG-9R enables transvascular delivery of siRNA to the central nervous system. a, Mice were injected intravenously with FITC-siRNA/peptide complexes, and uptake by brain, spleen and liver cells was examined by flow cytometry. Representative histograms (top) and cumulative data (bottom) are shown. Black, RV-MAT-9R; red, RVG-9R. Error bars indicate s.d. (*n* = 4). Three asterisks, *P* = 0.001. **b**, Coronal sections of brain from FITC-siRNA/RVG-9R-injected mice (*n* = 6) were stained with anti-FITC antibody and examined by fluorescence microscopy. Images of FITC-positive cells in the cortex, striatum and thalamus at lower magnifications (left panel) and higher magnifications of the boxed regions (middle panel) are shown. Right panel, images from control immunoglobulin-stained brain sections at the higher magnification. Scale bar, 200 μm.

complex was also found to be non-toxic in a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2*H*-tetrazolium bromide (MTT) assay (more than 90% viability at 48 h after treatment of Neuro 2a cells with RVG-9R at up to 25:1 peptide:siRNA ratio; data not shown).

RVG-9R enables transvascular siRNA delivery to the brain

For potential delivery *in vivo*, we first examined whether RVG-9R binding protects the siRNA against degradation from serum nucleases. Unlike naked siRNA, RVG-9R-bound siRNA was at least partly stable for up to 8 h (Supplementary Fig. S2). Next we tested whether RVG-9R could transport siRNA to brain cells *in vivo*. Mice were injected intravenously with FITC-siRNA complexed to RVG-9R or RV-MAT-9R; after 10 h, single-cell suspensions from the brain, spleen and liver were examined by flow cytometry. As shown in Fig. 3a, FITC fluorescence was detected in the brain only when the siRNA was complexed to RVG-9R. However, no FITC uptake was seen in the spleen or liver, suggesting that RVG-9R allows specific targeting of brain cells. The presence of FITC-positive cells in different regions

throughout the mouse brain was also confirmed by microscopic examination of brain sections stained with anti-FITC antibody (Fig. 3b).

To test brain-specific gene silencing, GFP transgenic mice were injected intravenously with GFP siRNA bound to RVG-9R or RV-MAT-9R on three consecutive days; their brain, spleen and liver cells were examined for GFP expression 2 days after the last injection. GFP expression was much greater in the brain than in the spleen and liver in the transgenic mice. Despite this, a significant decrease in GFP expression was seen after treatment with RVG-9R-bound siRNA but not with RV-MAT-9R-bound siRNA (Fig. 4a). Moreover, GFP silencing was seen only in the brain and not in the liver or spleen, confirming the specificity of brain targeting. To confirm these results in a different system, we also targeted an endogenous gene. Wild-type Balb/c mice were injected intravenously with an siRNA targeting the mouse gene encoding Cu-Zn superoxide dismutase 1 (SOD1; ref. 19) complexed to RVG-9R or RV-MAT-9R, and mRNA and protein levels of SOD1 in the brain, spleen and liver were measured by quantitative polymerase chain reaction (PCR) and western blotting, respectively. Although no changes in SOD1 levels were detected in any organ in RV-MAT-9R/siRNA-treated animals, both messenger RNA and protein levels of SOD1 were significantly decreased in the brain, but not in other organs, in the RVG-9R/siRNA-treated mice (Fig. 4b).

To confirm that the observed knockdown was due to specific delivery of siRNA within the brain, we also tested for the presence of *SOD1* siRNA by northern blot analysis. siRNA was detected in the brain but not in the spleen or liver of treated mice (Fig. 4c). Both the gene silencing effect and siRNA detectability in the brain cells gradually decreased over a 9-day period (Fig. 4d and data not shown), in agreement with the duration of silencing reported after local administration of siRNA in the brain[20]. Repeated administration of RVG-9R/siRNA complex neither induced inflammatory cytokines nor elicited an anti-peptide antibody response (Supplementary Fig. S3), attesting to the viability of this delivery approach. Taken together, our results show that RVG-9R enables the intravenous delivery of siRNA to silence gene expression within the brain.

RVG-9R/siRNA treatment for viral encephalitis

We have reported that intracranial treatment with antiviral siRNAs can robustly protect mice from fatal flaviviral encephalitis[8]. However, a noninvasive intravenous treatment method would be optimal for

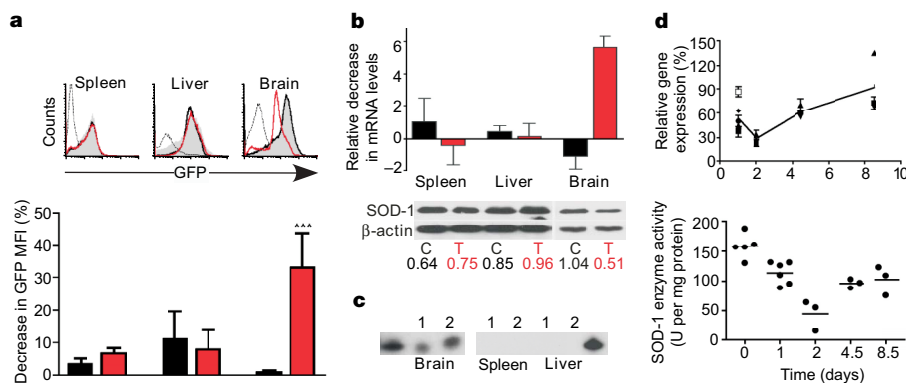**Figure 4 | Brain-specific gene silencing by intravenous injection of RVG-9R/siRNA complex. a**, GFP transgenic mice were injected intravenously with GFP siRNA/peptide complexes, and their brain, spleen and liver cells were analysed for GFP expression. Representative histograms (top) and cumulative data (bottom) are shown. Error bars indicate s.d. (*n* = 5); asterisks, *P* = 0.004. Dotted lines in the upper panel, cells from wild-type mice; grey fill, mock; black lines and columns, RV-MAT-9R; red lines and columns, RVG-9R. **b**, Balb/c mice were injected intravenously with *SOD1* siRNA/peptide complexes, and their brain, spleen and livers were examined for *SOD1* mRNA (top) and SOD1 protein levels (bottom). Black columns, RV-MAT-9R (C); red columns,

RVG-9R (T). Error bars indicate s.d. (*n* = 3). The numbers below the western blot represent the ratios of band intensities of SOD-1 normalized to that of β-actin. **c**, Small RNAs isolated from different organs of RVG-9R/*SOD1* siRNA-injected mice were probed with siRNA sense strand oligonucleotide. Antisense strand oligonucleotide was used as positive control (first and last lanes). **d**, Mice were injected intravenously with *SOD* siRNA bound to RVG-9R, and the duration of gene silencing was determined by quantification of *SOD1* mRNA levels (top) and SOD1 protein enzyme activity (bottom) on the indicated days after siRNA administration. Error bars in the upper panel indicate s.d. (*n* = 4); the horizontal lines in the lower panel represent mean values.
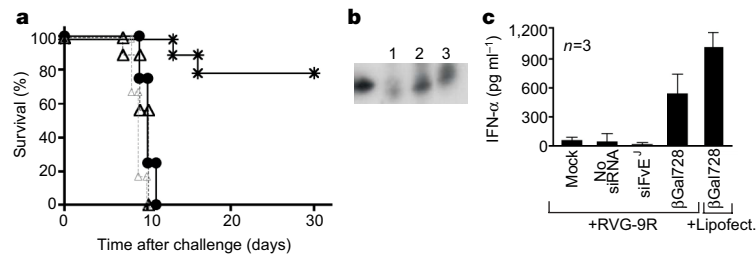
Figure 5 | Intravenous treatment with antiviral siRNA/RVG-9R complex protects mice against JEV encephalitis. a, JEV-infected mice were treated intravenously with siLuc or siFvE^J complexed to either RVG-9R or RV-MAT-9R daily for 4 days and monitored for survival. Grey triangles, mock (no siRNA); black triangles, RVG-9R plus siLuc; circles, RV-MAT-9R plus siFvE^J; stars, RVG-9R plus siFvE^J. $n = 9$. **b**, RNA isolated from the brains of RVG-9R/siFvE^J-treated mice were probed with siFvE^J sense strand to examine for the presence of siRNA antisense strand by northern blotting. Antisense strand of siFvE^J served as a positive control (first lane). **c**, Balb/c mice were injected intravenously with siFvE^J bound to RVG-9R peptide, and 7 h later their serum samples were tested for IFN-α levels by enzyme-linked immunosorbent assay. The immunostimulatory βgal 728 siRNA complexed with RVG-9R or Lipofectamine was used as a positive control. Error bars indicate s.d.

clinical use. We therefore tested whether intravenous treatment with siRNA bound to RVG-9R could protect mice from JEV-induced encephalitis. Unlike wild-type mice, immunodeficient mice are uniformly susceptible to peripheral infection with flaviviruses[21,22]. We therefore infected NOD/SCID mice with JEV ($5\,LD_{50}$) intraperitoneally followed 4 h later by intravenous treatment with antiviral FvE^J siRNA (siFvE^J)[8] or an irrelevant luciferase siRNA (siLuc), complexed with RVG-9R or RV-MAT-9R. The siRNA treatment was repeated on three successive days and the mice were observed for survival for at least 30 days. Untreated mice and mice treated with siFvE^J complexed to RV-MAT-9R or with siLuc complexed to RVG-9R all died within 10 days, showing that neither the chimaeric peptides by themselves nor the irrelevant siRNA bound to RVG-9R affected the course of the disease. In contrast, treatment with siFvE^J complexed to RVG-9R resulted in about 80% survival (Fig. 5a). The presence of siFvE^J siRNA in the brains was also confirmed by northern blot analysis (Fig. 5b). To rule out the possibility that non-specific interferon (IFN) production mediated the protection observed, we measured serum IFN levels after administration of RVG-9R/siFvE^J. Although IFN levels were higher when mice were treated with a known immunostimulatory siRNA[23], IFN was not induced in RVG-9R/siFvE^J-treated animals (Fig. 5c), suggesting that the protection was mediated by RNA interference. Thus, intravenous treatment with RVG-9R/siRNA can be used for the treatment of viral encephalitis.

Discussion

Taken together, our results suggest that RVG-9R peptide may enable transvascular delivery of siRNA to the central nervous system. The relatively modest (about 50%) knockdown obtained is comparable to that reported after prolonged infusion of siRNA in the central nervous system[24,25]. However, many aspects of this delivery system could be refined to enhance the delivery efficacy. For instance, because RVG-9R-bound siRNA was only partly protected against degradation in the serum (Supplementary Fig. S2), the use of chemically stabilized siRNA[26] may enhance the efficacy of delivery. Moreover, encapsulation of even a stabilized siRNA within a liposomal nanoparticle greatly enhances serum half-life and bioavailability[27,28], and liposomal and polymeric nanopraticles coated with targeting ligands have been used for delivery in previous studies[29–31]. Thus, a combination of these methods to generate stabilized siRNA-encapsulated nanoparticles, coated with RVG peptide as a targeting ligand, may provide an ideal method to enhance delivery and decrease the requirement for siRNA and peptide for effective gene silencing. Moreover, RVG-coated nanoparticles may also provide a method for targeted brain delivery of other gene therapy vectors and small-molecule drugs. Direct conjugation of siRNA to the peptide[32] might be an alternative strategy to improve delivery.

Further studies to localize the presence of siRNA and gene silencing in different cell types within the brain are also needed to understand the mechanism by which RVG-9R enables delivery to the brain.

However, because RVG peptide alone (without 9R) was also detectable in the brain after intravenous injection (Fig. 1e), it is likely that receptor-mediated transcytosis by means of the α7 subunit of the AchR (which is widely expressed in the brain, including by capillary endothelial cells[13]) is involved in the process. The fact that RVG-9R, but not RV-MAT-9R, facilitated crossing of the BBB also indicates that specific receptor binding might be important. Although cell-penetrating peptides might also enable covalently conjugated cargo to cross membranes[33,34], receptor clustering mediated by a 1:10 molar ratio of siRNA/RVG-9R binding may be required (particularly when the siRNA is non-covalently bound to the peptide) for efficient transport of siRNA to neuronal cells. This may explain the neuronal cell specificity of targeting by RVG-9R. Because RVG-9R-delivered siRNA was functional in gene silencing in multiple systems, siRNA seems to detach from the peptide inside the cell, although how exactly this happens is unclear. Similarly, siRNA complexed with protamine has been reported to be effective in gene silencing[35]. Thus, although further studies are needed to explain the exact mechanisms, our study highlights the potential of RVG-9R to mediate transvascular delivery of siRNAs to the central nervous system. RVG-mediated delivery might also allow the use of RNA interference for the systematic analysis of gene function in brain cells under experimental settings. In principle, RVG-assisted delivery might also be used for the brain-directed transport of other therapeutic molecules such as gene therapy vectors and small-molecule drugs.

METHODS SUMMARY

For peptide binding studies, cells were incubated for 20 min with biotinylated peptides, washed and then stained with streptavidin–phycoerythrin (SAPE). For all siRNA delivery studies, siRNA was incubated with peptides at a 1:10 molar ratio for 10–15 min at room temperature ($20\,^{\circ}C$) in serum-free DMEM medium (for *in vitro* studies) or 5% glucose (for *in vivo* studies) before use. For all *in vivo* delivery experiments, mice were injected into the tail vein with siRNA/peptide complexes in 100–200 μl of 5% glucose, and the mice received 50 μg of siRNA in each injection. All statistical analyses comparing groups of mice treated with test and control peptides were performed by one-way analysis of variance followed by Bonferroni's post hoc test. $P < 0.05$ was considered significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. de Boer, A. G. & Gaillard, P. J. Drug Targeting to the brain. *Annu. Rev. Pharmacol. Toxicol.* **47**, 323–355 (2007).
2. Miller, G. Drug targeting. Breaking down barriers. *Science* **297**, 1116–1118 (2002).
3. Schlachetzki, F., Zhang, Y., Boado, R. J. & Pardridge, W. M. Gene therapy of the brain: the trans-vascular approach. *Neurology* **62**, 1275–1281 (2004).
4. Lentz, T. L., Burrage, T. G., Smith, A. L., Crick, J. & Tignor, G. H. Is the acetylcholine receptor a rabies virus receptor? *Science* **215**, 182–184 (1982).
5. Lafon, M. Rabies virus receptors. *J. Neurovirol.* **11**, 82–87 (2005).

6. Rubinson, D. A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genet.* **33**, 401–406 (2003).

7. Mazarakis, N. D. *et al.* Rabies virus glycoprotein pseudotyping of lentiviral vectors enables retrograde axonal transport and access to the nervous system after peripheral delivery. *Hum. Mol. Genet.* **10**, 2109–2121 (2001).

8. Kumar, P., Lee, S. K., Shankar, P. & Manjunath, N. A single siRNA suppresses fatal encephalitis induced by two different flaviviruses. *PLoS Med.* **3**, e96 (2006).

9. Leonard, S. & Bertrand, D. Neuronal nicotinic receptors: from structure to function. *Nicotine Tob. Res.* **3**, 203–223 (2001).

10. Lentz, T. L. Rabies virus binding to an acetylcholine receptor alpha-subunit peptide. *J. Mol. Recognit.* **3**, 82–88 (1990).

11. Notter, M. F. & Leary, J. F. Flow cytometric analysis of tetanus toxin binding to neuroblastoma cells. *J. Cell. Physiol.* **125**, 476–484 (1985).

12. Chen, T. J., Chen, S. S., Wu, R. E., Wang, D. C. & Lin, C. H. Implication of nNOS in the enlargement of AChR aggregates but not the initial aggregate formation in a novel coculture model. *Chin. J. Physiol.* **48**, 129–138 (2005).

13. Gotti, C. & Clementi, F. Neuronal nicotinic receptors: from structure to pathology. *Prog. Neurobiol.* **74**, 363–396 (2004).

14. Gupta, B., Levchenko, T. S. & Torchilin, V. P. Intracellular delivery of large molecules and small particles by cell-penetrating proteins and peptides. *Adv. Drug Deliv. Rev.* **57**, 637–651 (2005).

15. Dietz, G. P. & Bahr, M. Peptide-enhanced cellular internalization of proteins in neuroscience. *Brain Res. Bull.* **68**, 103–114 (2005).

16. Deshayes, S., Morris, M. C., Divita, G. & Heitz, F. Cell-penetrating peptides: tools for intracellular delivery of therapeutics. *Cell. Mol. Life Sci.* **62**, 1839–1849 (2005).

17. Wender, P. A. *et al.* The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. *Proc. Natl Acad. Sci. USA* **97**, 13003–13008 (2000).

18. Kim, W. J. *et al.* Cholesteryl oligoarginine delivering vascular endothelial growth factor siRNA effectively inhibits tumor growth in colon adenocarcinoma. *Mol. Ther.* **14**, 343–350 (2006).

19. Hino, T. *et al.* In vivo delivery of small interfering RNA targeting brain capillary endothelial cells. *Biochem. Biophys. Res. Commun.* **340**, 263–267 (2006).

20. Akaneya, Y., Jiang, B. & Tsumoto, T. RNAi-induced gene silencing by local electroporation in targeting brain region. *J. Neurophysiol.* **93**, 594–602 (2005).

21. Diamond, M. S., Shrestha, B., Marri, A., Mahan, D. & Engle, M. B cells and antibody play critical roles in the immediate defense of disseminated infection by West Nile encephalitis virus. *J. Virol.* **77**, 2578–2586 (2003).

22. Neyts, J., Leyssen, P. & De Clercq, E. Infections with flaviviridae. *Verh. K. Acad. Geneeskd Belg.* **61**, 661–697 (1999).

23. Judge, A. D. *et al.* Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nature Biotechnol.* **23**, 457–462 (2005).

24. Hoyer, D. *et al.* Global down-regulation of gene expression in the brain using RNA interference, with emphasis on monoamine transporters and GPCRs: implications for target characterization in psychiatric and neurological disorders. *J. Recept. Signal Transduct. Res.* **26**, 527–547 (2006).

25. Dorn, G. *et al.* siRNA relieves chronic neuropathic pain. *Nucleic Acids Res.* **32**, e49 (2004).

26. Soutschek, J. *et al.* Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432**, 173–178 (2004).

27. Morrissey, D. V. *et al.* Potent and persistent *in vivo* anti-HBV activity of chemically modified siRNAs. *Nature Biotechnol.* **23**, 1002–1007 (2005).

28. Zimmermann, T. S. *et al.* RNAi-mediated gene silencing in non-human primates. *Nature* **441**, 111–114 (2006).

29. Noble, C. O. *et al.* Development of ligand-targeted liposomes for cancer therapy. *Expert Opin. Ther. Targets* **8**, 335–353 (2004).

30. Zhang, Y. *et al.* Intravenous RNA interference gene therapy targeting the human epidermal growth factor receptor prolongs survival in intracranial brain cancer. *Clin. Cancer Res.* **10**, 3667–3677 (2004).

31. Hu-Lieskovan, S., Heidel, J. D., Bartlett, D. W., Davis, M. E. & Triche, T. J. Sequence-specific knockdown of EWS-FLI1 by targeted, nonviral delivery of small interfering RNA inhibits tumor growth in a murine model of metastatic Ewing's sarcoma. *Cancer Res.* **65**, 8984–8992 (2005).

32. Juliano, R. L. Peptide-oligonucleotide conjugates for the delivery of antisense and siRNA. *Curr. Opin. Mol. Ther.* **7**, 132–136 (2005).

33. Melikov, K. & Chernomordik, L. V. Arginine-rich cell penetrating peptides: from endosomal uptake to nuclear delivery. *Cell. Mol. Life Sci.* **62**, 2739–2749 (2005).

34. Henriques, S. T., Melo, M. N. & Castanho, M. A. Cell-penetrating peptides and antimicrobial peptides: how different are they? *Biochem. J.* **399**, 1–7 (2006).

35. Song, E. *et al.* Antibody mediated *in vivo* delivery of small interfering RNAs via cell-surface receptors. *Nature Biotechnol.* **23**, 709–717 (2005).

36. Schutz, B. *et al.* The oral antidiabetic pioglitazone protects from neurodegeneration and amyotrophic lateral sclerosis-like symptoms in superoxide dismutase-G93A transgenic mice. *J. Neurosci.* **25**, 7805–7812 (2005).

METHODS

Lentiviral experiments. shFvE[J] and shLuc lentiviral constructs and experiments to test protection against intracranial JEV challenge in mice have been reported previously[8]. Lentiviruses pseudotyped with VSV-G or RVG were generated by transfection of the lentiviral vector plasmids together with the helper plasmid pHR′8.9ΔVPR (core protein) and either the pVSV-G or pLTR-RVG envelope constructs into 293T cells. Culture supernatants were harvested after 48 h, and viral particles were concentrated by ultracentrifugation. Lentiviruses were spin-infected onto Neuro 2a or HeLa cells in the presence of Polybrene, and after 48 h the transduction efficiency was determined by analysing GFP expression by flow cytometry.

Peptides and siRNAs. Peptides RVG (YTIWMPENPRPGTPCDIFTNSRGK-RASNG), RV-MAT (MNLLRKIVKNRRDEDTQKSSPASAPLDDG), RVG-9R (YTIWMPENPRPGTPCDIFTNSRGKRASNGGGGRRRRRRRRR) and RV-MAT-9R (MNLLRKIVKNRRDEDTQKSSPASAPLDGGGGRRRRRRRRR) were synthesized and purified by high-performance liquid chromatography at the Tufts University Core Facility. RVG and RV-MAT peptides were also biotinylated at the carboxy terminus. In RVG-9R and RV-MAT 9R peptides, the C-terminal nine arginine residues were D-arginine.

siRNAs used in the studies included those targeting GFP (siGFP), firefly luciferase (siLuc) described in ref. 8, the envelope gene of JEV (siFvE[J]) and those targeting murine Cu-Zn superoxide dismutase (SOD-1)[19], and β-galactosidase (βGal728) bearing a motif eliciting interferon production[23]. For some experiments, siRNA with FITC label at the 3′ end of the sense strand was used. siRNAs were synthesized at Samchully Pharm. Co. Ltd or obtained from Dharmacon, Inc.

Peptide binding assay. For peptide binding studies, Neuro 2a, HeLa, CHO, 293T and BHK21 cell lines and single-cell suspensions made from freshly isolated mouse brain, spleen or liver were used. Cells were incubated in PBS with 2.5 μM biotinylated peptides for 20 min at 4 °C, washed three times with PBS and then treated with SAPE (BD Pharmingen) before analysis by flow cytometry. For competition experiments, cells were incubated with 2.5 μM biotinylated RVG peptide in the absence or presence of different concentrations of BTX (Sigma).

EMSA. For gel mobility-shift assays, 100 pmol of siRNA was incubated with peptides at 10:1, 1:1 and 1:10 molar ratios of siRNA to peptide for 15 min, subjected to electrophoresis on 2% agarose gels and stained with ethidium bromide. siRNA without peptide or incubated with biotinylated RVG (without 9R) served as controls.

Cytotoxicity assay. To test the cytotoxicity of RVG-9R/siRNA complexes, Neuro 2a cells (triplicates of 2×10^5 cells per well in 12-well plates) were incubated with different concentrations of peptide/siRNA complexes for 24–48 h before viability was determined with a standard MTT assay.

siRNA transduction and gene silencing *in vitro*. Uptake of siRNA into cells was monitored with FITC-labelled siLuc. siRNA (100 pmol) was incubated with different concentrations of RVG-9R or RV-MAT-9R in serum-free DMEM for 15 min at room temperature. The complexes were then added to Neuro 2a and HeLa cell cultures (plated at 5×10^4 cells per well in 12-well plates on the previous day). After incubation for 4 h at 37 °C the medium was replaced with 2 ml of fresh medium supplemented with 10% fetal bovine serum (Invitrogen) and the cells were cultured for a further 8–10 h before being examined by flow cytometry. Transfection with Lipofectamine 2000 was performed in accordance with the manufacturer's instructions.

To test gene silencing, Neuro 2a cells stably expressing GFP after transduction with the pLL3.7 lentiviral vector were incubated with 100 pmol of siGFP complexed with peptides at a 10:1 peptide/siRNA ratio and GFP expression analysed 48 h after transduction.

Animal experiments for testing siRNA delivery and gene silencing. Balb/c, C57BL/6-Tg(ACTB-EGFP)1Osb/J and NOD/SCID mice were purchased from Jackson Laboratories and used at 4–6 weeks of age. All mouse experiments had been approved by the CBR Institute (CBRI) institutional review board, and animal infection experiments were performed in a biosafety level 3 animal facility at the CBRI.

To test peptide uptake by brain cells, 200 μg of biotinylated peptides in 0.2 ml of PBS were injected into tail veins of Balb/c mice; 4 h later, single-cell suspensions of brains were permeabilized, treated with SAPE and analysed by flow cytometry. For all siRNA delivery experiments, peptide/siRNA complexes (at a peptide to siRNA molar ratio of 10:1) were prepared in 100–200 μl of 5% glucose and injected intravenously at 50 μg of siRNA per mouse per injection. To test FITC-siRNA delivery, Balb/c mice were injected twice with siRNA, 6 h apart, and organs were harvested after a further 10 h. To test GFP silencing, C57BL/6-Tg(ACTB-EGFP)1Osb/J mice were injected with peptide/siRNA complexes on three consecutive days and organs were harvested 2 days later. For SOD-1 silencing, Balb/c mice were given three injections of siRNA/peptides at 8-h intervals and organs were harvested at various time points. For testing protection against

JEV encephalitis, NOD/SCID mice were challenged intraperitoneally with 5 LD$_{50}$ of JEV (LD$_{50}$, the lethal dose for half of the mice, was predetermined by using serial dilutions of the virus) 4 h before treatment with intravenous peptide/siRNA was started. The siRNA treatment was repeated at 24-h intervals for a total of 4 days.

Staining of brain sections. Mice were injected twice with RVG-9R-bound siRNA-FITC; brains were harvested 10–12 h later. Brains were sectioned frozen on a sliding microtome to a thickness of 40 μm and incubated for 48 h at 4 °C with mouse anti-FITC antibodies (20 μg ml^{-1}; Jackson Immuno Research) or isotype controls (IgG1 from murine myeloma; 20 μg ml^{-1}; Sigma). Sections were washed and FITC immunoreactivity was detected with Alexa-488 goat anti-mouse secondary antibodies (dilution 1:500; Invitrogen). The antibody enhancement was performed because only one in six mice revealed FITC-positive cells by direct examination.

Quantitative RT–PCR. Total RNA was isolated from different organs of peptide/SOD1 siRNA-treated mice by using an RNeasy RNA isolation kit (Qiagen). The RNA was reverse transcribed with Superscript III and random hexamers (Invitrogen) in accordance with the manufacturer's protocol. Real-time PCR was performed on 1 μl of complementary DNA, or on a comparable amount of RNA without reverse transcriptase, with the QuantiTect SYBR Green PCR kit (Qiagen) in accordance with the manufacturer's instructions. Amplification conditions were as follows: 40 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 30 s with a Bio-Rad iCycler. Primers for GAPDH and SOD-1 have been described previously[36]. Specificity was verified by melt-curve analysis and agarose-gel electrophoresis. SOD-1 mRNA levels from the test animals, normalized with glyceraldehyde-3-phosphate dehydrogenase levels, were divided by the equivalent values from untreated mice to calculate relative changes.

Western blot analysis. Cell suspensions from mouse tissue were homogenized in buffer containing 25 mM HEPES pH 7.5, 300 mM NaCl, 1.5 mM MgCl$_2$, 0.1% Triton X-100, 0.2 mM EDTA and 0.5 mM dithiothreitol and protease-inhibitor cocktail (Complete-Mini; Roche Diagnostic). The samples (10 μg of protein each) were subjected to electrophoresis on 15% SDS–polyacrylamide mini gels (Bio-Rad) and transferred to a poly(vinylidene difluoride) membrane. The membrane was probed with anti-β-actin antibodies (Sigma) or anti-SOD1 antibodies (Stressgen Biotechnologies) and detected with an enhanced chemiluminescence western blot system (Pierce Biotechnologies). The blots were scanned and the ratio of band intensities of SOD-1 normalized to β-actin was calculated with Image J software.

Determination of SOD1 enzyme activity. The level of Cu/Zn SOD-1 enzyme activity in brain tissue was measured with the SOD-1 Assay Kit-WST (Cell Technologies, Inc.) in accordance with the manufacturer's instructions. Frozen brain tissues were homogenized in ice-cold sucrose buffer (50 mM sucrose, 200 mM mannitol, 1 mM EDTA in 10 mM Tris-HCl buffer pH 7.4) and used in ELISA after inactivation of Mn/Fe SOD-1 with chloroform/ethanol. The enzyme activity is denoted as units per milligram of total protein in the brain tissue.

Northern blot to detect siRNA. Small RNA (5 μg) extracted from cell suspensions with an miRNeasy mini kit (Qiagen) were subjected to electrophoresis on 15% TBE-urea PAGE gels (Invitrogen), transferred to a positively charged nylon membrane (BrightStar-plus; Ambion) and probed with sense siRNA probes as described previously[8]. Antisense strand of synthetic SOD-1 siRNA (200 fmol) (first and last lanes in Fig. 4c) was used as a positive control.

IFN response. Balb/c mice were injected intravenously with 50 μg of either siFvE[J] or siβGal728 complexed with RVG-9R peptide. siβGal728 complexed to Lipofectamine-2000 served as a positive control. Serum samples obtained 7 h after siRNA treatment were tested for IFN-α levels with a mouse type-I IFN detection ELISA kit (PBL Biomedical Laboratories), in accordance with the manufacturer's instructions.

Serum stability. Naked and RVG-9R-complexed siRNA (100 pmol) were incubated at 37 °C in 50% FBS (Invitrogen), 90% human serum or 90% mouse serum. Aliquots taken at different time points were treated with Proteinase K and frozen in $2 \times$ urea TBE-loading buffer. All samples were subjected to electrophoresis on 15% TBE-urea polyacrylamide gels under non-denaturing conditions and detected by staining with SYBR gold.

Immunogenicity studies. Balb/c mice were injected intravenously with 50 μg of siLuc complexed to RVG-9R peptide or, for positive control, with 25 μg of trinitrophenol-conjugated keyhole-limpet haemocyanin-biotin (TNP-KLH-biotin) peptide (Biosearch Technologies). The injection was repeated on days 3, 10 and 22, and serum samples were collected on days 21 and 30. To detect the presence of mouse antibodies to RVG or biotin (in the TNP-KLH-biotin-injected mice), 1:10 and 1:50 dilutions of sera were incubated in 96-well microtitre plates coated with biotinylated RVG peptide (1 μg per well). The presence of binding mouse antibodies was detected with anti-mouse immunoglobulins conjugated to

horseradish peroxidase. A mouse cytokine/chemokine array kit (Ray Biotech Inc.) was used to detect a panel of 56 secreted cytokines and chemokines in the serum obtained 1 day after the fourth injection of RVG-9R/siRNA. For positive control, mice were injected with 50 µg of lipopolysaccharide and serum was collected after 15 h. The manufacturer's recommended protocol was used to perform the assay.

Quantification and statistical analysis. Western blots and cytokine array profiles were quantified by determining band intensities with ImageJ public domain software from the National Institutes of Health (http://rsb.info.nih.gov/ij/).

ARTICLES

CD1d–lipid-antigen recognition by the semi-invariant NKT T-cell receptor

Natalie A. Borg[1], Kwok S. Wun[1], Lars Kjer-Nielsen[2], Matthew C. J. Wilce[1], Daniel G. Pellicci[2], Ruide Koh[1], Gurdyal S. Besra[3], Mandvi Bharadwaj[2], Dale I. Godfrey[2], James McCluskey[2] & Jamie Rossjohn[1]

The CD1 family is a large cluster of non-polymorphic, major histocompatibility complex (MHC) class-I-like molecules that bind distinct lipid-based antigens that are recognized by T cells. The most studied group of T cells that interact with lipid antigens are natural killer T (NKT) cells, which characteristically express a semi-invariant T-cell receptor (NKT TCR) that specifically recognizes the CD1 family member, CD1d. NKT-cell-mediated recognition of the CD1d–antigen complex has been implicated in microbial immunity, tumour immunity, autoimmunity and allergy. Here we describe the structure of a human NKT TCR in complex with CD1d bound to the potent NKT-cell agonist α-galactosylceramide, the archetypal CD1d-restricted glycolipid. In contrast to T-cell receptor–peptide-antigen–MHC complexes, the NKT TCR docked parallel to, and at the extreme end of the CD1d-binding cleft, which enables a lock-and-key type interaction with the lipid antigen. The structure provides a basis for the interaction between the highly conserved NKT TCR α-chain and the CD1d–antigen complex that is typified in innate immunity, and also indicates how variability of the NKT TCR β-chain can impact on recognition of other CD1d–antigen complexes. These findings provide direct insight into how a T-cell receptor recognizes a lipid-antigen-presenting molecule of the immune system.

Polymorphic major histocompatibility complex (MHC) molecules present peptide antigens that are specifically co-recognized by αβ T-cell receptors (TCRs) expressed on T cells. These TCRs are highly specific and restricted to recognizing the MHC molecules of the individual from which they were derived; this concept is known as MHC-restriction[1]. In contrast, CD1 family members are mono-morphic MHC class-I-like glycoproteins that present lipid-based antigens recognized by T cells from different individuals and even across species. The antigen-binding cleft of the CD1 family contains large hydrophobic pockets that are ideally suited to binding lipid antigens and, on the basis of sequence identity and chromosomal location, are divided into CD1a, CD1b, CD1c, CD1d and CD1e family members[2]. Although humans express all family members, only the *CD1d* genes are present and expressed in mice and rats. CD1a, b and c molecules, which can present self and foreign (microbial) lipid antigens, are generally recognized by polyclonal T cells expressing αβ TCRs or by some γδ T cells that recognize CD1c. Despite considerable data on TCR–peptide-MHC (pMHC) interactions[3,4], the mechanism of TCR interaction with CD1 molecules is not understood.

NKT cells are uniquely restricted to CD1d proteins presenting self or foreign lipid antigens. NKT cells typically express a semi-invariant TCR (NKT TCR; see Methods) that comprises an invariant α-chain, and a restricted TCR-β-chain repertoire (reviewed in ref. 5). NKT cells are implicated in a broad range of diseases (reviewed in ref. 6): they can enhance the response to some bacterial, viral and parasitic infections, and some types of cancer, yet can suppress autoimmune disease, allograft rejection and graft versus host disease. In some instances, they promote undesirable immune responses associated with allergy, some types of autoimmunity, and atherosclerosis[6]. Accordingly, NKT cells represent an important immunotherapeutic target with broad clinical potential. Although several CD1d-restricted glycolipid agonists for NKT cells have been described,

α-galactosylceramide (α-GalCer) is the prototypical NKT-cell antigen and is by far the most studied, with proven capacity to potently stimulate NKT cell activation in mice and humans. Indeed, α-GalCer has been the subject of several clinical trials for its anti-cancer potential[6].

A distinct footprint

To establish how CD1d–α-GalCer is specifically recognized by a human NKT TCR, we expressed and purified both components to form the NKT TCR–CD1d–α-GalCer complex (ref. 7 and Methods), which was then crystallized and the structure was determined to 3.2 Å resolution with an R_{fac} and R_{free} of 22.6 and 29.3%, respectively (see Methods and Supplementary Table 1). The initial experimental phases clearly showed unbiased electron density for the α-GalCer (Supplementary Fig. 1); moreover, the electron density at the NKT TCR–CD1d–α-GalCer interface was unambiguous (Supplementary Fig. 2). The crystals contained two nearly indistinguishable ternary complexes (root mean square deviation 0.514 Å over 817 Cα atoms) within the asymmetric unit, so structural analyses will be confined to one NKT TCR–CD1d–α-GalCer complex.

The NKT TCR adopted an acute docking mode, positioned over the extreme end, directly above the F′ pocket of the CD1d-antigen binding cleft, spanning residues 76–89 and 147–151 of the α1-helix and the α2-helix, respectively (Fig. 1a, c, e). This extreme NKT TCR docking mode has not been observed in any of the TCR–pMHC complexes determined to date and is completely different to the hypothetical models of the NKT TCR–CD1d ternary complex that were previously generated[7,8]. Moreover the NKT TCR bound approximately parallel to the long axis of the CD1d-antigen-binding cleft, which is distinct from the broad range of 'diagonal' footprints observed for MHC class-I-restricted TCRs (Fig. 1b, d, f). Furthermore, of the three MHC class I positions (65, 69 and 155) that are postulated to represent conserved docking points guiding

[1]The Protein Crystallography Unit, ARC Centre of Excellence in Structural and Functional Microbial Genomics, Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria 3800, Australia. [2]Department of Microbiology & Immunology, University of Melbourne, Parkville, Victoria 3010, Australia. [3]School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

MHC-restriction[9], no equivalent positions are contacted in the CD1d-restricted interaction (Fig. 1d). Accordingly, the salient features of MHC–peptide-mediated recognition by a TCR are not preserved in CD1d–lipid-mediated recognition.

NKT TCR–CD1d interaction

The human NKT TCR uses an invariant α-chain and CDR3α loop (Vα24–Jα18, where V is 'variable' and J is 'joining') combined with a Vβ11-containing β-chain, although there are marked differences in the length and composition of CDR3β sequences characterized from NKT cells[10,11]. The total buried surface area (BSA) on binding between the NKT TCR and CD1d is only ≈ 910 Å^2, which falls below the range of BSA values (1,200–2,400 Å^2) observed at the TCR–pMHC interface[4](Fig. 1e). Within this small NKT TCR–CD1d–α-GalCer

interface, the α-chain contributes more per cent of the BSA than the β-chain (65.5% versus 34.5% respectively), which is consistent with the α-chain dominating contacts with CD1d–α-GalCer in comparison to the β-chain (approximately 82 versus 32 contacts, respectively; Supplementary Table 2). This marked skewing of the interactions explains the basis of the selection of the semi-invariant NKT TCR. The Vβ11 chain usage seems to be dictated by the CDR2β loop and associated framework residues (Tyr 48β and Glu 56β), because the CDR1β loop does not mediate any contacts, and the role of the CDR3β loop (6% of the BSA) in this ternary complex is limited to Tyr 103β forming van der Waals contacts with Gln 150 of the α2-helix of CD1d (Supplementary Table 2). In contrast, the CDR2β loop, which contributed 27.5% of the BSA, formed a focused stretch of predominantly polar interactions exclusively with the α1-helix

Figure 1 | Comparison of TCR recognition of CD1d-lipid and pMHC-I. Overview of the NKT TCR in complex with human CD1d–α-GalCer and in comparison to a typical TCR (LC13) interaction with pMHC-I (HLA-B8-FLR)[36]. **a**, NKT TCR–human CD1d–α-GalCer complex. Human CD1d heterodimer, green; α-GalCer, magenta; NKT TCR α-chain and β-chain, yellow and blue respectively. **b**, The TCR–pMHC-I complex (PDB accession code 1MI5). MHC, grey; peptide, magenta; TCR α-chain and β-chain, purple and cyan respectively. **c**, Parallel docking mode of NKT TCR CDR loops onto

human CD1d–α-GalCer. Human CD1d helices, pale green; α-GalCer, magenta; CDR loops: CDR1α, yellow; CDR2α, green; CDR3α, cyan; CDR1β, magenta; CDR2β, orange; CDR3β, blue. **d**, Diagonal docking mode of a typical TCR interaction with pMHC-I. CDR loops are coloured as above. For **c** and **d**, the centre of mass between the Vα and Vβ domain is shown as a black line. **e**, Footprint of the NKT TCR on the surface of human CD1d–α-GalCer. **f**, Footprint of LC13 on the surface of HLA-B8-FLR. For **e** and **f**, colour-coding as in **c**.

(residues 83–89) of CD1d (Supplementary Table 2). Specifically, Tyr 48β and Tyr 50β converge to form 3 H-bonds with Glu 83; Tyr 48β also H-bonds to Lys 86, which in turn forms a salt bridge with Glu 56β. Asn 53β, at the tip of the CDR2β loop, also formed van der Waals contacts with Arg 89 (Fig. 2a). Interestingly, the mouse–human NKT TCR–CD1d cross-reactivity correlates with the high level of sequence similarity in the CDR2β loops of mouse Vβ8.2 and human Vβ11 (Supplementary Fig. 3). In contrast, rat NKT cells are restricted to their syngeneic CD1d–α-GalCer and this restriction is linked to the CDR2β sequence in the rat NKT TCR[12], further emphasizing the important role of CDR2β in contacting CD1d.

The invariant NKT TCR α-chain interactions are mediated only by the CDR3α and CDR1α loops (52.2% and 10.5% of the BSA, respectively) (Figs 1, 2 and Supplementary Table 2). The CDR1α loop interacts solely with the α-GalCer, whereas the CDR3α loop straddles the antigen-binding cleft, interacting with the α1-helix, the α2-helix and α-GalCer. The CDR3α interactions are mediated exclusively through the Jα18-region, which is consistent with the lack of NKT cells in TCR Jα18-gene-targeted mice[13], and why some CD1d-restricted human NKT cells preserve Jα18 usage despite diverging from Vα24 expression[14]. The CDR3α-mediated interactions were largely electrostatic in nature, but also included some van der Waals-mediated contacts, including Leu 99α, which plugged a small cavity between Phe 84 and Val 147; and Ser 97α, on the tip of the CDR3α loop, which interacted with Gln 150 (Fig. 2b). There was an unusual inter-digitation of arginine residues at the CDR3α–CD1d interface, in which the guanidinium group of Arg 79 from CD1d was flanked, in an anti-parallel manner, with Arg 103α and Arg 95α. This cluster of positively charged residues was dissipated by neighbouring acidic groups,

including Asp 94α, which interacted with Arg 79, Arg 95α and Arg 103α, and Asp 80 that salt-bridged to Arg 95α, the latter of which also formed a H-bond with Ser 76 (Fig. 2b). Additionally, Thr 98α H-bonded to Gln 150 and, consequently, all Jα18 residues at the tip of the CDR3α loop mediated contacts with CD1d–α-GalCer (Supplementary Table 2). Accordingly, the CDR2β loop and the CDR3α loop seem to have prominent roles in enabling a CD1d-restricted response. Consistent with our findings, several of the CD1d contact points identified here (Arg 79, Asp 80, Glu 83 and Asp 151) have previously been identified in mouse CD1d mutagenesis studies as being important in α-GalCer-mediated NKT cell activation[15–17]. Notably, of the interactions made with CD1d, Ser 76, Asp 80, Glu 83, Lys 86, Met 87 and Asp 151 are unique to CD1d (Supplementary Fig. 4) in comparison to the other CD1 family members that interact with T-cells, thereby providing insight into why the NKT TCR is restricted to CD1d.

Lipid recognition

α-GalCer, the therapeutic candidate and possibly the microbial mimetic of naturally occurring microbial α-glycuronosylceramides[18], is a potent agonist for NKT cells and considered a prototypic ligand for the NKT TCR when presented by CD1d. Consistent with the previously determined structures[19,20], α-GalCer protrudes minimally from the CD1d cleft with only the glycosyl head group exposed for recognition by the NKT TCR, interacting solely with the CDR1α and CDR3α loops (with each loop contributing roughly equivalent BSA contributions) (Fig. 3 and Supplementary Table 2). The galactose ring sits below the CDR1α loop and abuts the CDR3α loop, being sandwiched between Trp 153 of CD1d and the aliphatic moiety of
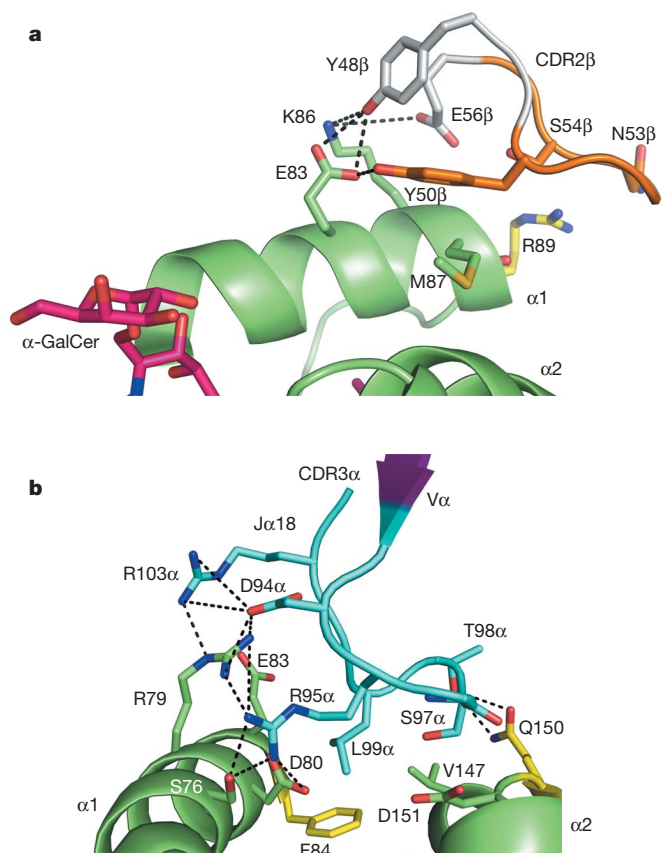

Figure 2 | CD1d-mediated interactions. NKT TCR contacts with CD1d are dominated by the CDR2β and CDR3α loops. **a**, CDR2β-mediated contacts. CDR2β, orange; CDR2β framework residues, white. **b**, CDR3α-mediated contacts. Vα gene segment of CDR3α, purple; Jα gene segment of CDR3α loop, cyan; conserved residues between mouse and human CD1d, green; non-conserved residues between mouse and human CD1d, yellow.
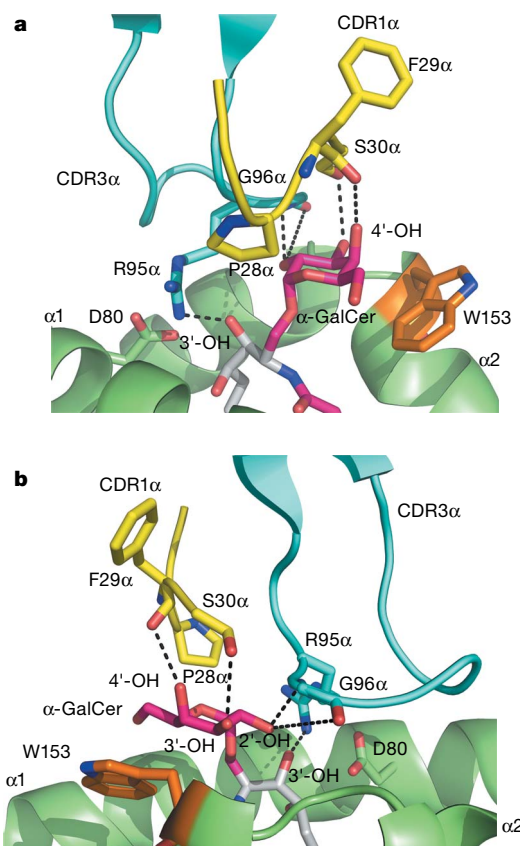

Figure 3 | α-GalCer-mediated interactions. NKT TCR contacts with α-GalCer are mediated by the CDR1α and CDR3α loops, shown in two orthogonal orientations (**a**, **b**). CDR1α, yellow; CDR3α, cyan; conserved residues between mouse and human CD1d, green; non-conserved residues between mouse and human CD1d, orange. α-GalCer galactose moiety and acyl tail, magenta; sphingosine tail, white.

Arg 95α, the side chain of which also H-bonds to the 3′ hydroxyl of the sphingosine chain (Fig. 3). The 2′, 3′ and 4′ hydroxyl groups of the galactose ring make H-bonds with the invariant TCR α-chain, with the 2′ hydroxyl and 4′ hydroxyl H-bonding to the main chain of Gly 96α and Phe 29α, respectively, whereas the 3′ hydroxyl H-bonds to Ser 30α. This mode of binding is consistent with the fine specificity the NKT TCR exhibits for α-GalCer and its closely related analogues, where, for example, CD1d–α-ManCer—which differs from α-GalCer in the positioning of the 2′ and 4′ hydroxyls—does not bind the NKT TCR, presumably owing to the loss of two H-bonds[21]. The preferred binding of the α-linked versus β-linked mono-galactose moiety is also evident because the latter is predicted to engender a more perpendicular orientation that would result in a disruption of CDR1α-mediated contacts. Thus the orientation of the α-linked galactose head group, which is largely sequestered by polar contacts, provides a basis for the profound stimulatory capacity of this potent agonist for the NKT TCR.

Species cross-reactivity

Human NKT TCRs cross-react with mouse CD1d–α-GalCer complexes (and vice versa) with approximately comparable affinity[7], highlighting the evolutionarily conserved nature of this interaction. Of the interactions mediated by CD1d, only three (positions 84, 89 and 150) are non-conserved between human and mouse (Fig. 2, Supplementary Table 2 and Supplementary Fig. 5), indicating that the human NKT TCR is likely to adopt a very similar footprint on the mouse CD1d–α-GalCer complex. Although the positioning of the galactose ring differs subtly between these human CD1d– and mouse CD1d–α-GalCer structures[22], our human NKT TCR–mouse CD1d–α-GalCer model (data not shown) suggests that the majority of α-GalCer contacts (in particular the H-bond mediated contacts) are preserved. Moreover, in comparing the sequence of the human NKT TCR to the homologous mouse NKT TCR (Vα14–Jα18, Vβ8.2), the residues that have an important role in recognizing CD1d–α-GalCer (CDR1α, CDR3α and CDR2β) are conserved across the mouse and human NKT TCR[7] (Supplementary Fig. 3). Interestingly though, the mouse Vβ7 and Vβ2 sequences do not show conservation to human Vβ11 within the CDR2β loop (Supplementary Fig. 3), and mouse NKT cells expressing these alternate TCR Vβ chains exhibit different affinities for mouse and human CD1d–α-GalCer[23,24]. Nevertheless, the 'conserved species hotspots' in the CD1d and the homologous NKT TCRs highlight the innate characteristics of the monomorphic CD1d interacting with a semi-invariant TCR; and this contrasts with the TCR recognition of MHC class-I molecules, which is highly restricted to syngeneic MHCs.

Lock-and-key interaction

Plasticity of the TCR is a hallmark of adaptive immunity in recognizing the pMHC, although plasticity of the peptide can also be notable[25]. We, and others, have previously determined the structure of the unliganded NKT TCR[7,8], and thus we were able to address whether NKT TCR malleability was important to the CD1d–α-GalCer interaction. Surprisingly, however, the NKT TCR did not change conformation appreciably on ligation, apart from a small alteration in the juxtapositioning of the Vα/Vβ domains and small movements in the tip of the CDR3α loop and in the inherently flexible CDR3β loop (not shown)[7]. There is some re-orientation of side-chain conformations on complexation that maximize contacts at the interface including Arg 79, Glu 83, Lys 86, and Arg 89 from CD1d, and Tyr50β and Tyr103β from the NKT TCR. Thus the CDR loops that contact with the CD1d–α-GalCer complex minimally change conformation on ligation. This predominantly rigid 'lock and key' interaction, largely attributable to the intramolecular interactions between the CDR loops in the unliganded NKT TCR, correlates with the lack of temperature dependence of the interactions[26] and with the affinity for the NKT TCR–CD1d–α-GalCer interaction ($K_d \approx 1$–10 μM) being at the higher end of the range

observed for TCRs interacting with pMHC ligands ($K_d \approx 1$–50 μM)[7,8], particularly evident for the mouse NKT TCR ($K_d \approx 0.1$–0.3 μM)[17,26].

Recognition of other CD1d-restricted ligands

Nevertheless, plasticity and/or variability of the NKT TCR is potentially a key feature in recognizing diverse CD1d-restricted antigens, such as isoglobotrihexosylceramide[27] (iGb3), α-glycuronosylceramides[28–30], α-galactosyldiacylglycerol[31], a tumour-derived disialoganglioside (GD3)[32], phosphatidylinositolmannoside (PIM) 4 (ref. 33), and phosphoethanolamine[34]. None of these antigens activates mouse NKT cells to the same extent as α-GalCer, which is probably due to lower affinity binding to the TCR and/or CD1d, combined with limited antigenicity stimulating only a subset of α-GalCer-reactive NKT cells. The orientation and position of the galactose ring with respect to the NKT TCR provides insight into this diversity of recognition (Fig. 4). First, the 6′ hydroxyl of the galactose ring points towards solvent that would enable accommodation of the 6′ carboxy substitution present in α-galacturonosylceramide (Fig. 4). Second, immediately adjacent to the 3′ and 4′ moieties (that mediate the common 3′ and 4′ glycosidic linkages) of the glycosyl head group is a large positively charged pocket, which is lined by the CDR3β loop (Fig. 4). The peripherally located CDR3β loop not only provides an opportunity to accommodate diverse CDR3β sequences but is also consistent with potentially engendering plasticity to accommodate different antigens (Fig. 4). For example, the more bulky sugar head
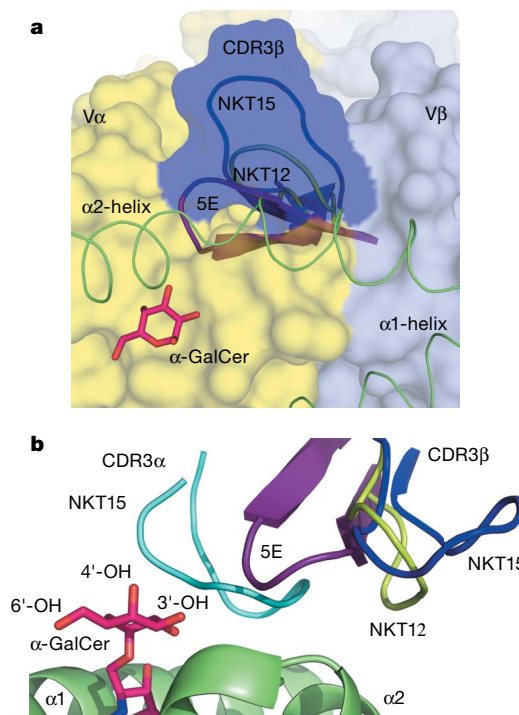
Figure 4 | CDR3β loop variability in NKT TCRs. Conformations of the CDR3β loop with respect to the α-GalCer moiety. **a**, Surface representation of the NKT15 T-cell receptor. The docking of NKT12 and the 5E TCR (PDB accession codes 2EYR and 2CDF, respectively) onto CD1d–α-GalCer was modelled by superposition from the NKT15–human CD1d–α-GalCer structure. The helices of CD1d are shown in green and the galactose moiety of α-GalCer is shown in magenta ball-and-stick format. The Vα and Vβ domains of NKT15 are shown in pale yellow and blue, respectively, and the surface representation of the CDR3β loop of NKT15 is coloured blue. The CDR3β loop of NKT12 and the 5E TCR are shown in lime green and purple, respectively. **b**, The NKT15 cavity lined by residues from the CDR3β loop (blue) and adjacent to the 3′ and 4′-hydroxyl moieties of the glycosyl head group of α-GalCer (magenta). The CDR3α loop of NKT15 is coloured cyan. The CDR3β loops of the NKT12 and 5E TCRs are coloured lime green and purple respectively. Human CD1d helices, green.

groups, such as iGb3, could be accommodated within this cavity and the flexible CDR3β loop. Consistent with this notion, NKT cells recognize iGB3 (α1-3-linked terminal galactose), but not Gb3 (α1-4-linked terminal galactose), indicating that the orientation of the terminal sugar moiety plays a critical part in TCR binding. Accordingly, the high degree of variability in the CDR3β sequences in NKT cells, coupled with the location of this loop within the complex (Fig. 4), is likely to be associated with mediating NKT cell subset specificity for different CD1d-antigens.

Discussion

NKT cells seem to be at the crossroads of innate and adaptive immunity. Their cross-species reactivity is indicative of the strong conservation of innate immune function, reflecting its ancient nature. On the other hand, the ability of NKT cells (or subsets thereof) to interact with distinct CD1d-antigen complexes highlights the adaptability of the limited NKT TCR repertoire. The structure reported here provides an opportunity to rationally design agonists and antagonists to modulate NKT-cell responses and provides insights into the seemingly hybrid roles of the NKT TCR. The invariant α-chain is critical in engendering the innate function of the receptor, whereas the adaptability of the interaction is potentially modulated by plasticity and/or variability of the CDR3β loop. It is unclear whether other NKT TCRs that exhibit differences in either Vα usage or differing Vβ sequences adopt a different docking strategy in recognizing CD1d, or, indeed, whether T-cells that are restricted to other CD1 family members also deviate from the docking mode observed here. Nevertheless, the structure simultaneously reveals a marked divergence of how TCRs interact with CD1d in comparison to pMHC complexes and provides the first atomic insight into lipid recognition by a TCR.

METHODS

Protein expression, refolding and purification. The cloning, expression and purification of the human NKT (NKT15) TCR used in this study (α-chain: TRAV10-TRAJ18; β-chain: TRBV25-1,TRBD1 and TRBJ2-7, with the CDR3β sequence and flanking residues being CASS GLRDRGL YEQYFGPGTRLTVT) and human CD1d have been previously described[7]. Briefly, the NKT TCR was expressed in *Escherichia coli*, purified from inclusion bodies and refolded prior to purification. Soluble human CD1d was expressed using HI5 insect cells, purified using Ni-sepharose beads (GE Healthcare) and buffer-exchanged into 10 mM Tris pH 8.0, 150 mM NaCl. The endogenous ligand was displaced with a 3:1 molar ratio of α-GalCer (sourced from G. Besra and Alexis Biochemicals) to protein at room temperature overnight. Excess α-GalCer was removed from human CD1d-α-GalCer using a Superdex 200 (10/30) column. Purified human CD1d–α-GalCer and the NKT TCR were then incubated together at a 1:1 ratio and concentrated to 8 mg ml^{-1} in 10 mM Tris pH 8.0, 150 mM NaCl for crystallization.

Crystallization, structure determination and refinement. Plate-like crystals of NKT TCR–human CD1d–α-GalCer complex were grown using the hanging drop vapour diffusion technique. The crystal grew at room temperature in a 1:1 ratio of protein to mother liquor containing 13% PEG 10K, 0.1 M bis-tris propane, pH 8.5, 0.2 M tri-sodium citrate. The crystals, which diffracted to 3.2 Å resolution, belong to space group *C*2, with two complex molecules per asymmetric unit (Supplementary Table 1). Crystals were flash-frozen prior to data collection using stepwise increments of glycerol.

The crystal structure was determined by the molecular replacement method, using the program Molrep[35] within the CCP4i suite. The previously determined structures of the unliganded NKT TCR (termed NKT15; PDB accession code 2EYS) and the human 3.0 Å CD1d–α-GalCer (with α-GalCer removed to prevent model bias) were used as the search model. Unbiased features in the electron density map confirmed the correctness of the molecular replacement solution. Although the structure was solved at moderate resolution, the quality of the electron density was excellent (Supplementary Figs 1 and 2), and, as the structures of the unliganded complexes were previously determined, the structure refined very well for 2 ternary complexes in the asymmetric unit at this resolution (Supplementary Table 1). CCP4i was used for iterative rounds of refinement and the progress of refinement was monitored by the R_{free} value. Initially, rigid-body refinement was used and subsequently restrained refinement using translation libration screw parameters and non-crystallographic symmetry restraints.

1. Zinkernagel, R. M. & Doherty, P. C. Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**, 701–702 (1974).
2. Brigl, M. & Brenner, M. B. CD1: Antigen Presentation and T Cell Function. *Annu. Rev. Immunol.* **22**, 817–890 (2004).
3. Clements, C. S. *et al.* Specificity on a knife-edge: the αβ T cell receptor. *Curr. Opin. Struct. Biol.* **16**, 787–795 (2006).
4. Rudolph, M. G. *et al.* How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).
5. Godfrey, D. I. *et al.* NKT cells: what's in a name? *Nature Rev. Immunol.* **4**, 231–237 (2004).
6. Godfrey, D. I. & Kronenberg, M. Going both ways: immune regulation via CD1d-dependent NKT cells. *J. Clin. Invest.* **114**, 1379–1388 (2004).
7. Kjer-Nielsen, L. *et al.* A structural basis for selection and cross-species reactivity of the semi-invariant NKT cell receptor in CD1d/glycolipid recognition. *J. Exp. Med.* **203**, 661–673 (2006).
8. Gadola, S. D. *et al.* Structure and binding kinetics of three different human CD1d α-galactosylceramide-specific T cell receptors. *J. Exp. Med.* **203**, 699–710 (2006).
9. Tynan, F. E. *et al.* T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. *Nature Immunol.* **6**, 1114–1122 (2005).
10. Dellabona, P. *et al.* An invariant V alpha 24-J alpha Q/V beta 11 T cell receptor is expressed in all individuals by clonally expanded CD4-8- cells. *J. Exp. Med.* **180**, 1171–1176 (1994).
11. Porcelli, S. *et al.* Analysis of T cell antigen receptor (TCR) expression by human peripheral blood CD4−8− alpha/beta T cells demonstrates preferential use of several V beta genes and an invariant TCR alpha chain. *J. Exp. Med.* **178**, 1–16 (1993).
12. Pyz, E. *et al.* The complementarity determining region 2 of BV8S2 (V beta 8.2) contributes to antigen recognition by rat invariant NKT cell TCR. *J. Immunol.* **176**, 7447–7455 (2006).
13. Cui, J. Q. *et al.* Requirement for Vα14 NKT cells in IL-12-mediated rejection of tumors. *Science* **278**, 1623–1626 (1997).
14. Gadola, S. D. *et al.* Vα24-JαQ-independent, CD1d-restricted recognition of α-galactosylceramide by human CD4[+] and CD8 αβ[+] T lymphocytes. *J. Immunol.* **168**, 5514–5520 (2002).
15. Burdin, N. *et al.* Structural requirements for antigen presentation by mouse CD1. *Proc. Natl Acad. Sci. USA* **97**, 10156–10161 (2000).
16. Kamada, N. *et al.* Crucial amino acid residues of mouse CD1d for glycolipid ligand presentation to Vα14 NKT cells. *Int. Immunol.* **13**, 853–861 (2001).
17. Sidobre, S. *et al.* The Vα14 NKT cell TCR exhibits high-affinity binding to a glycolipid/CD1d complex. *J. Immunol.* **169**, 1340–1348 (2002).
18. Bendelac, A. *et al.* The biology of NKT cells. *Annu. Rev. Immunol.* **25**, 297–336 (2007).
19. Koch, M. *et al.* The crystal structure of human CD1d with and without alpha-galactosylceramide. *Nature Immunol.* **6**, 819–826 (2005).
20. Zajonc, D. M. *et al.* Structure and function of a potent agonist for the semi-invariant natural killer T cell receptor. *Nature Immunol.* **6**, 810–818 (2005).
21. Sidobre, S. *et al.* The T cell antigen receptor expressed by Vα14i NKT cells has a unique mode of glycosphingolipid antigen recognition. *Proc. Natl Acad. Sci. USA* **101**, 12254–12259 (2004).
22. Godfrey, D. I. *et al.* CD1d antigen presentation: treats for NKT cells. *Nature Immunol.* **6**, 754–756 (2005).
23. Schumann, J. *et al.* Cutting edge: influence of the TCR Vβ domain on the avidity of CD1d:α-galactosylceramide binding by invariant Vα14 NKT Cells. *J. Immunol.* **170**, 5815–5819 (2003).
24. Schumann, J. *et al.* Targeted expression of human CD1d in transgenic mice reveals independent roles for thymocytes and thymic APCs in positive and negative selection of Vα14i NKT Cells. *J. Immunol.* **175**, 7303–7310 (2005).
25. Tynan, F. E. *et al.* A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. *Nature Immunol.* **8**, 268–276 (2007).
26. Cantu, C. III *et al.* The paradox of immune molecular recognition of α-galactosylceramide: low affinity, low specificity for CD1d, high affinity for αβ TCRs. *J. Immunol.* **170**, 4673–4682 (2003).
27. Zhou, D. *et al.* Lysosomal glycosphingolipid recognition by NKT cells. *Science* **306**, 1786–1789 (2004).
28. Kinjo, Y. *et al.* Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* **434**, 520–525 (2005).
29. Mattner, J. *et al.* Exogenous and endogenous glycolipid antigens activate NKT cells during microbial infections. *Nature* **434**, 525–529 (2005).
30. Sriram, V. *et al.* Cell wall glycosphingolipids of *Sphingomonas paucimobilis* are CD1d-specific ligands for NKT cells. *Eur. J. Immunol.* **35**, 1692–1701 (2005).
31. Kinjo, Y. *et al.* Natural killer T cells recognize diacylglycerol antigens from pathogenic bacteria. *Nature Immunol.* **7**, 978–986 (2006).
32. Wu, D. Y. *et al.* Cross-presentation of disialoganglioside GD3 to natural killer T cells. *J. Exp. Med.* **198**, 173–181 (2003).

33. Fischer, K. *et al.* Mycobacterial phosphatidylinositol mannoside is a natural antigen for CD1d-restricted T cells. *Proc. Natl Acad. Sci. USA* **101**, 10685–10690 (2004).

34. Rauch, J. *et al.* Structural features of the acyl chain determine self-phospholipid antigen recognition by a CD1d-restricted invariant NKT (iNKT) cell. *J. Biol. Chem.* **278**, 47508–47515 (2003).

35. Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr.* **56**, 1622–1624 (2000).

36. Kjer-Nielsen, L. *et al.* A structural basis for the selection of dominant αβ T cell receptors in antiviral immunity. *Immunity* **18**, 53–64 (2003).

Author Contributions N.A.B. performed experimental work, analysis and prepared the manuscript; K.S.W., L.K.N., M.C.J.W., D.P., M.B. performed experimental work; G.S.B. provided essential reagents; D.I.G. devised the project, analysed data and wrote the manuscript; and J.M. and J.R. are joint senior and corresponding authors—they co-led the investigation, devised the project, analysed the data and wrote the manuscript.

Author Information The X-ray crystallographic coordinates and structure factor files have been deposited in the Protein Data Bank with accession number 2PO6. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R. (jamie.rossjohn@med.monash.edu.au) or J.M. (jamesm1@unimelb.edu.au).

LETTERS

Hyperion's sponge-like appearance

P. C. Thomas[1], J. W. Armstrong[2], S. W. Asmar[2], J. A. Burns[1], T. Denk[3], B. Giese[4], P. Helfenstein[1], L. Iess[5], T. V. Johnson[2], A. McEwen[6], L. Nicolaisen[7], C. Porco[8], N. Rappaport[2], J. Richardson[1], L. Somenzi[5], P. Tortora[9], E. P. Turtle[10] & J. Veverka[1]

Hyperion is Saturn's largest known irregularly shaped satellite and the only moon observed to undergo chaotic rotation[1–3]. Previous work has identified Hyperion's surface as distinct from other small icy objects[4,5] but left the causes unsettled. Here we report high-resolution images that reveal a unique sponge-like appearance at scales of a few kilometres. Mapping shows a high surface density of relatively well-preserved craters two to ten kilometres across. We have also determined Hyperion's size and mass, and calculated the mean density as $544 \pm 50 \, \mathrm{kg \, m^{-3}}$, which indicates a porosity of >40 per cent. The high porosity may enhance preservation of craters by minimizing the amount of ejecta produced or retained[6,7], and accordingly may be the crucial factor in crafting this unusual surface.

Hyperion orbits Saturn every 21.3 days with a semi-major axis of 24.55 Saturn radii. The Cassini spacecraft flew by Hyperion four times in 2005 and 2006 at ranges less than 300,000 km. A close fly-by in September 2005 was specifically targeted[8] to obtain high-resolution remote-sensing data and to track directly changes in the spacecraft trajectory caused by the mass of Hyperion.

Imaging during 2005 confirmed the expectations[1–3,9,10] of chaotic rotation, and showed that the object remains in an average spin state similar to that observed by Voyager: rotating nearly about the long axis, at ~72°–75° per day (see Supplementary Information). Cassini observations demonstrate that the spin vector moves through the body and across the sky; the subsolar latitude varied by 70° over the year of observations. The movement of the spin vector through the body meant that a global shape model (Fig. 1) could be constructed only by merging several partial models determined for each of the different encounters. The resulting model has a mean radius of 135 ± 4 km, with radii from the centre of figure between 95 and 182 km (Table 1). The uncertainty in the mean radius, hence of the volume, derives largely from errors in combining the partial shape models.

Hyperion's mass was obtained from measurements of the space-craft velocity during the close encounter on 25 September 2006. Measurements entailed the transmission from the ground of a highly stable, almost monochromatic signal at 7.2 GHz and its coherent retransmission to Earth at 8.4 GHz and 32.5 GHz. Significant measurement noise, about 8–10 times larger than in previous Cassini radio science experiments[11] was due to the small solar elongation angle (55°) and the lack of a Ka-band uplink carrier.

The spacecraft fly-by took place at a closest approach distance of 618 km, and a relative velocity of $5.6 \, \mathrm{km \, s^{-1}}$. The resultant change in the spacecraft velocity was about $0.1 \, \mathrm{m \, s^{-1}}$, much larger than the measurement noise. The observed Doppler data were fitted with a dynamical model that included the gravitational acceleration from all Saturn system bodies and the non-gravitational accelerations. Thirteen parameters were fitted, namely the components of Cassini and Hyperion state vectors at a reference epoch, and the satellite's GM value, where G is the gravitational constant and M the mass.

A stable orbital solution was obtained resulting in a GM of $0.375 \pm 0.003 \, \mathrm{km^3 \, s^{-2}}$. The corresponding mass M, assuming $G = 6.6742 \times 10^{-11} \, \mathrm{m^3 \, kg^{-1} \, s^{-2}}$, is therefore $5.619 \pm 0.050 \times 10^{18}$ kg. This value is in agreement with an independent estimate derived by the Cassini navigation team using a different method of orbit determination[8].

Hyperion has a mean density of $544 \pm 50 \, \mathrm{kg \, m^{-3}}$; the uncertainty derives largely from the volume measurement. The mean density indicates a porosity of $42 \pm 6\%$ if Hyperion is primarily water ice[12]. Any significant contributions from higher density components increase the porosity. Hyperion's density is similar to that obtained for smaller satellites orbiting near the rings (360–$690 \, \mathrm{kg \, m^{-3}}$; 15–89 km mean radii[13]), but only one-third that of Phoebe[14] ($1,630 \, \mathrm{kg \, m^{-3}}$), the one other small object imaged by Cassini at similar resolutions and phase angles.

The most striking visual aspect of Hyperion is its sponge-like appearance (Fig. 1a, h) which is unlike that of any other object imaged to date and was not suggested[4,5] in the low-resolution Voyager images, although Hyperion had an intermediate albedo, spotted colour distribution, and topography that differed from those of other small icy satellites[4,5]. We measured the sizes of impact features: depressions that have distinct rim morphology, cut other forms and are approximately circular. After tabulating the obvious craters, few depressions remained, and we conclude that the sponge-like appearance is primarily manifested in the density of well-preserved craters 2–10 km in diameter. The area density of Hyperion craters exceeding 10 km in size, near 'empirical saturation'[15], is similar to that on Phoebe (Fig. 2a). On both objects the crater density decreases for craters between 10 km and 1 km in diameter, which is probably a reflection of a production function[14] with 'cookie cutter' removal of small craters by larger ones. For craters with diameters between 2 km and 11 km, Hyperion has a cumulative number of craters per unit area twice that on Phoebe; at similar resolution these objects are strikingly different (Fig. 1g, h).

Enhanced preservation of craters in this size range on Hyperion is probably the key to Hyperion's appearance. Craters may be eroded or otherwise rendered uncountable either through covering by ejecta, direct erasure by other craters, erosion by very small craters, or other effects such as sublimation. Hyperion displays little evidence of ejecta, and there is positive evidence that the retained ejecta volume

[1]Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA. [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. [3]Institut für Geologische Wissenschaften, Freie Universität, 12249 Berlin, Germany. [4]Institute of Planetary Research, German Aerospace Center, Rutherfordstrasse 2, 12489 Berlin, Germany. [5]Dipartimento di Ingegneria Aerospaziale ed Astronautica, Università La Sapienza, via Eudossiana 18, 00184 Roma, Italy. [6]Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA. [7]Physics and Astronomy Department, University of California at Los Angeles, Box 951547 Los Angeles, California 90095, USA. [8]Space Science Institute, 4750 Walnut Street, Boulder, Colorado 80301, USA. [9]DIEM-II Facoltà di Ingegneria, Università di Bologna, Via Fontanelle, 4047100 Forlì, Italy. [10]Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA.
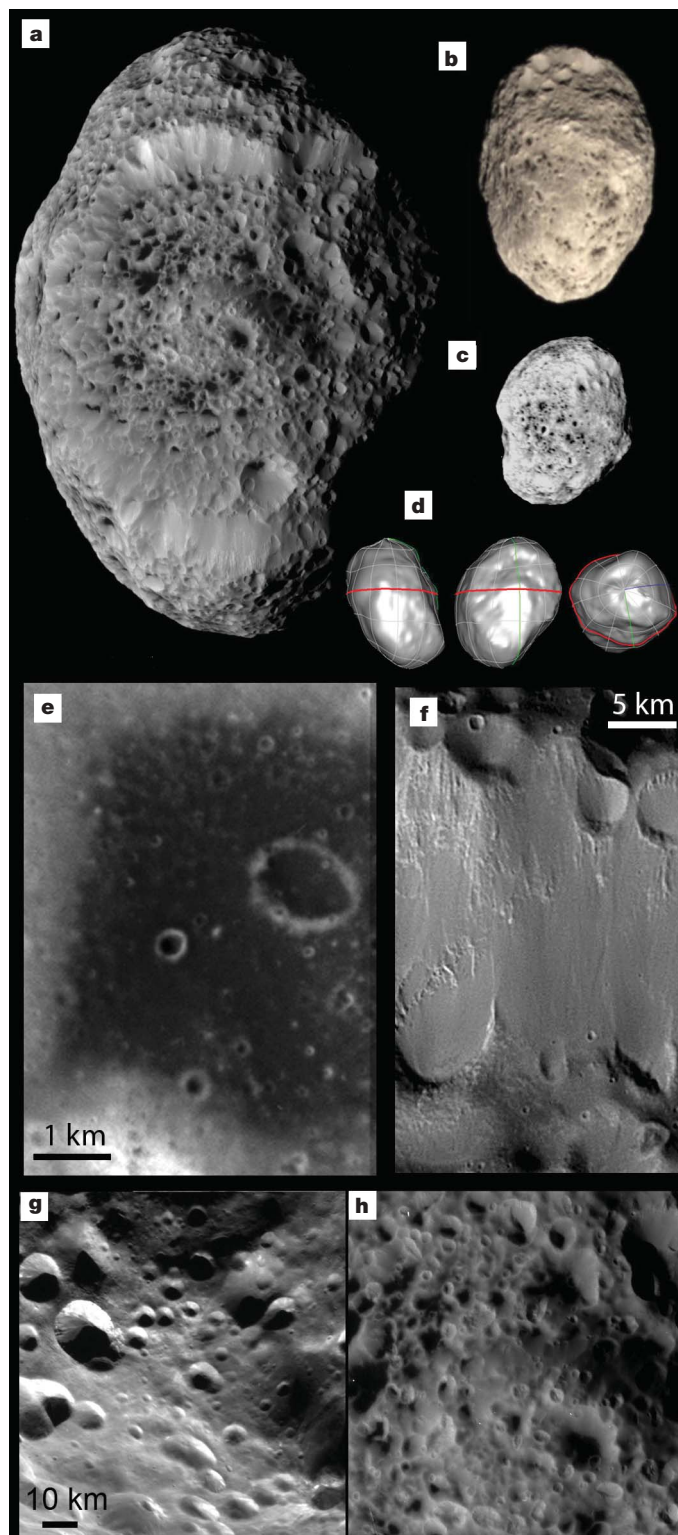
Table 1 | Properties of Hyperion

Property	Value
Mean radius	135 ± 4 km
Range of radii	95–182 km
Mass	$(5.619 \pm 0.050) \times 10^{18}$ kg
Mean density	544 ± 50 kg m^{-3}
Range of gravitational heights	0–43.8 km
Range of surface acceleration	1.7–2.1 cm s^{-2}

Gravitational heights are heights above an equipotential surface.

is small. Old crater floors are themselves heavily cratered (Fig. 1e), as shown by slightly brighter rims visible with strong contrast enhancement. The survival of these short rims (likely to have been less than 15 m in height[16,17]) shows that the large craters have experienced less than $\sim$20 m of deposition or erosion. The distinct visibility of narrow rim zones on craters, even ones cut by younger impacts, suggests the presence of structural rims unaffected by superposed ejecta. Ejecta blocks are scarce compared to Phoebe[14].

High porosity may contribute to the apparent minimal effects of ejecta either by suppressing production and dispersal outside the crater[6] or by enhanced velocity of material, such that most escapes the body[7]. Experiments and scaling theory predict that impacts into high-porosity targets produce less ejecta mass if the craters are primarily produced by compression rather than excavation[6]. A crucial part of these predictions is the role of gravity, parameterized by gD, where g is the local acceleration due to gravity, and D is the crater diameter. For a target of given porosity, the larger gD is, the smaller is the ratio of ejecta to crater volume. For a particular gD, at higher porosities the ejecta ratio is predicted to decrease partly by increased compression and partly by lower ejection velocities that allow more material to fall within the crater. Porosities below 30% have little effect on ejecta production, those over 40% can reduce ejecta by factors of more than four. Assuming the modelling results[6], Fig. 2b predicts that among observed targets Hyperion should have the most significant porosity effects on ejecta volumes. A different role for porosity has been suggested[7] wherein enhanced ejecta velocities cause escape from the target body. The only empirical information on cratering on a highly porous object comes from the Deep Impact experiment[18] on an object of radius 3 km with a porosity of probably >50%. Although it is unclear how large a fraction exceeded the escape velocity of $\sim$1.3 m s^{-1}, it is likely that at least a modest fraction was retained[18,19]. Hyperion's escape velocity ranges from 45 m s^{-1} to 99 m s^{-1} depending upon location, and thus it may be difficult for large fractions of any ejecta produced to escape.

It is unlikely that unusual crater depths significantly enhance the sponge-like appearance. Crater depth-to-diameter ratios for the 13 examples that can be measured reliably using shadow lengths average 0.21 ± 0.05 (s.d.). These ratios are similar to values for fresh lunar craters[16], and are slightly greater than for some small rocky objects[20]. They are slightly larger than that for craters on large icy satellites, 0.14 (ref. 21).

Adding to the unusual appearance of Hyperion are dark surfaces in the floors of degraded craters (Fig. 1). These have broadband-visible (0.56 µm) albedos a factor of $\sim$4 less than the average value of $\sim$0.5 for Hyperion. The dark materials are slightly redder than brighter areas: UV/GRN (0.338 µm/0.568 µm) = 0.61 for dark areas and 0.65 for bright slopes. IR3/GRN (0.930 µm/0.568 µm) ratios are $\sim$1.3 for bright areas; 1.6 for dark ones. These colours suggest mixing of brighter and darker materials broadly similar to that seen on some other dark Saturnian satellites[22]. Dark floors are largely absent from fresher craters and avoid regional slopes of >15°. Margins of the low-albedo areas are gradational; there are no exposed dark layers, no

Figure 1 | Surface features of Hyperion. a, Full disk view, 300 km top to bottom, showing the sponge-like appearance and distribution of darker material in craters. The image includes the largest probable impact feature. Image N15506383386, range = 62145 km, solar phase angle $\alpha = 51.5°$. **b**, Colour composite showing side opposite to that shown in **a** from images N1530199676, N1530199776 and N1530199809. **c**, Image N1497114664, showing partial profile of 120-km crater (left side). From 212000 km, $\alpha = 12.9°$. **d**, Orthogonal views of the shape model of Hyperion. The latitude–longitude grid is relative to the early Revolution 15 spin vector, UTC 2005, 268/04:29:06 to 268/17:20:16 (see Supplementary Information). **e**, Dark floor of old crater, showing large number of sub-kilometre craters. Contrast has been greatly enhanced by both high-pass filtering and stretching; brighter rims are dark compared to average Hyperion. Portion of image N1506393614. **f**, Slopes of the 240-km diameter crater, showing smaller crater (left side) elongated upslope, partial burial of other craters, and rugged upper rim. Portion of image N1506391600. **g**, Phoebe, image N1465669953, 0.195 km per pixel, $\alpha = 83.7°$. **h**, Hyperion, image N1506388580, 0.197 km per pixel, $\alpha = 51.1°$.

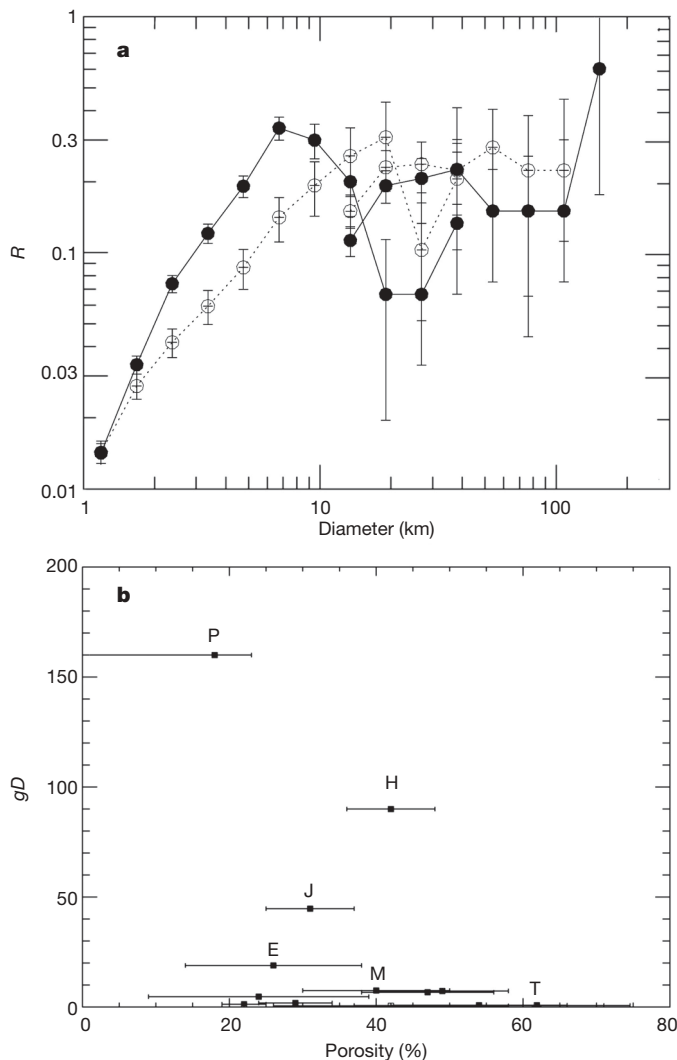**Figure 2 | Impact crater characteristics and conditions on Hyperion. a**, Area
density of craters on Hyperion and Phoebe. Solid symbols are Hyperion
data; open symbols are for Phoebe. Data are restricted to diameters >1 km;
resolution of data extending to 1 km is approximately 100 m per pixel. Error
bars are from $\sqrt{n}$ scaling, where n is the number of craters in each size bin.
$R = [(D_a D_b)3/2]^*[N_a/(D_b - D_a)]$ where D_a and D_b are crater diameter
ranges, $D_b > D_a$ and N_a is the number of craters per unit area. **b**, Porosity and
gravity parameter for well-imaged objects. gD is local gravity times crater
diameter of one-third the object mean radius (cm s^{-2} km). Porosity below
30% is predicted to have little effect on ejecta volume[6]. Comparison objects
are labelled with the first letters of Phoebe, Janus, Epimetheus, Mathilde, and
Tempel-1; others include Phobos, Deimos, Ida, Pan, Atlas, Prometheus and
Pandora. Porosities and gD calculated from information in refs 13, 14,
28–30. gD for Hyperion is 3,000 times that for Tempel-1. Error bars reflect
uncertainties in mean density apart from uncertainties in composition. For
Phoebe, the lower bound of 0% reflects uncertainty in composition.

scale. Any exogenous dark material, such as debris from Phoebe or
other outer satellites[23] would be evenly distributed initially and
subsequently concentrated even by small amounts of downslope
transport.

Could sublimation, enhanced by the dark material, contribute to
deepening and widening the craters[24] to give Hyperion's distinctive
appearance? Our measurements show that craters 2–10 km in dia-
meter have depths of >400 m. Deepening by sublimation would have
to be over 100 m, and widening would have to be several hundreds of
metres to make a visible morphological difference. Approximate
scaling from modelling of Ceres[25], Callisto[26] and comets[27] suggests
that at Saturn's distance from the Sun, dark water ice will sublimate
much less than 10 m over the period of existence of the Solar System
so far, even allowing for some concentration of incident radiation at
the bottom of depressions. CO_2 is far more volatile[26,27], and has been
detected on Hyperion[24]. However, the restriction of erosional effects
to <20 m may limit the amount of CO_2 available to sublimate. Thus,
although sublimation and downslope motion may affect the distri-
bution of dark material, primary crater morphology rather than sub-
limation dictates the unusual appearance of Hyperion.

1. Klavetter, J. J. Rotation of Hyperion. I—Observations. *Astron. J.* **97**, 570–579
    (1989a).
2. Klavetter, J. J. Rotation of Hyperion. II—Dynamics. *Astron. J.* **98**, 1855–1874
    (1989b).
3. Black, G. J., Nicholson, P. D. & Thomas, P. C. Hyperion: rotational dynamics. *Icarus*
    117, 149–161 (1995).
4. Smith, B. A. *et al.* A new look at the Saturn system: the Voyager 2 images. *Science*
    215, 504–537 (1982).
5. Thomas, P. C., Black, G. J. & Nicholson, P. D. Hyperion: Rotation, shape, and
    geology from Voyager images. *Icarus* **117**, 128–148 (1995).
6. Housen, K. R. & Holsapple, K. A. Impact cratering on porous asteroids. *Icarus* **163**,
    102–119 (2003).
7. Asphaug, E., Ryan, E. V. & Zuber, M. T. in *Asteroids III* (eds Bottke, W. F., Cellino, A.,
    Paolicchi, P. & Binzel, R. P.) 463–484 (Tucson, Univ. Arizona Press, 2002).
8. Criddle, K. E. *et al.* Cassini orbit determination: the only targeted Hyperion flyby of
    the prime mission. *29th Annual AAS Guidance and Control Conference
    (Breckenridge, Colorado)* AAS Paper~06–082 (American Astronomical Soceiety,
    San Diego, California, 2006).
9. Wisdom, J., Peale, S. J. & Mignard, F. The chaotic rotation of Hyperion. *Icarus* **58**,
    137–152 (1984).
10. Binzel, R. P., Green, J. R. & Opal, C. B. Chaotic rotation of Hyperion? *Nature* **320**, 511
    (1986).
11. Armstrong, J. W., Iess, L., Tortora, P. & Bertotti, B. Stochastic gravitational wave
    background: upper limits in the 10^{-6}-10^{-3} Hz band. *Astrophys. J.* **599**, 806–813
    (2003).
12. Clark, R. N., Fanale, F. P. & Gaffey, M. J. in *Satellites* (eds Burns, J. A. & Mathews, M.
    S.) 437–491 (Tucson, Univ. Arizona Press, 1986).
13. Porco, C. C. *et al.* Physical characteristics and possible accretionary origins for
    Saturn's small satellites. *Annu. Lunar Planet. Sci. Conf.* **37**, 2289 (2006).
14. Porco, C. C. *et al.* Cassini imaging science: initial results on Phoebe and Iapetus.
    Science **307**, 1237–1242 (2005).
15. Chapman, C. R. *et al.* Impact history of Eros: craters and boulders. *Icarus* **155**,
    104–118 (2002).
16. Pike, R. J. in *Impact and Explosion Cratering* (eds Roddy, D. J., Pepin, R. O. & Merrill,
    R. B.) 484–509 (Pergamon, New York, 1977).
17. Melosh, H. J. *Impact Cratering: A Geologic Process* 1–253 (Oxford Monographs on
    Geology and Geophysics, No. 11, Oxford Univ. Press, New York, 1989).
18. Richardson, J. E. & Melosh, H. J. Determining comet Tempel 1's gravity, mass, and
    density via the ballistic behavior of the Deep Impact ejecta plume. *Icarus*
    (submitted).
19. Sugita, S. *et al.* Subaru telescope observations of Deep Impact. *Science* **310**,
    274–278 (2005).
20. Carr, M. H. *et al.* The geology of Gaspra. *Icarus* **107**, 61–71 (1994).
21. Schenk, P. M. Ganymede and Callisto—Complex crater formation and planetary
    crusts. *J. Geophys. Res.* **96**, 15635–15664 (1991).
22. Buratti, B. J., Hicks, M. D. & Davies, A. Spectrophotometry of the small satellites of
    Saturn and their relationship to Iapetus, Phoebe, and Hyperion. *Icarus* **175**,
    490–495 (2005).
23. Cruikshank, D. P. *et al.* The dark side of Iapetus. *Icarus* **53**, 90–104 (1983).
24. Cruikshank, D. P. *et al.* Surface composition of Hyperion. *Nature* doi:10.1038/
    nature05948 (this issue).
25. Fanale, F. P. & Salvail, J. R. The water regime of asteroid (1) Ceres. *Icarus* **82**,
    97–110 (1989).
26. Moore, J. M. *et al.* Mass movement and landform degradation on the icy Galilean
    satellites: results of the Galileo nominal mission. *Icarus* **240**, 294–312 (1999).

morphologically distinct mounds of dark material, and no collec-
tions of dark material at the base of the large slopes in the largest
crater (Fig. 1f). The distribution and gradational margins of the dark
material are consistent with downslope motion of surficial materials.
Macroscopic compositional layering or other segregation of different
components in the 'bedrock' is not seen on Hyperion. There are no
bright-rayed craters, nor are there exposures of distinct brighter or
darker layers in Hyperion's crater walls as on Phoebe[14]. Downslope
movement of material other than that in the dark-floored depres-
sions is shown by a few crater-wall streamers, crater-filling on long
slopes, and some slumping inside craters. The lack of any clear
segregation of different materials suggests that if the dark material
is intrinsic to Hyperion it is mixed nearly uniformly at a small

27. Delsemme, A. H. in *Comets* (ed. Wilkening, L. L.) 85–130 (Tucson, Univ. Arizona Press, 1982).
28. Richardson, J. E., Melosh, H. J., Greenberg, R. J. & O'Brien, D. P. The global effects of impact-induced seismic activity on fractured asteroid surface morphology. *Icarus* **179**, 325–349 (2005).
29. Thomas, P. C. *et al.* Mathilde: size, shape, and geology. *Icarus* **140**, 17–27 (1999).
30. Thomas, P. C. Gravity, tides, and topography on small satellites and asteroids— Application to surface features of the Martian satellites. *Icarus* **105**, 326–244 (1993).

LETTERS

Surface composition of Hyperion

D. P. Cruikshank[1], J. B. Dalton[3], C. M. Dalle Ore[3], J. Bauer[4], K. Stephan[5], G. Filacchione[6], A. R. Hendrix[4], C. J. Hansen[4], A. Coradini[6], P. Cerroni[6], F. Tosi[6], F. Capaccioni[6], R. Jaumann[5], B. J. Buratti[4], R. N. Clark[8], R. H. Brown[9], R. M. Nelson[4], T. B. McCord[10], K. H. Baines[4], P. D. Nicholson[11], C. Sotin[12], A. W. Meyer[2], G. Bellucci[7], M. Combes[13], J.-P. Bibring[14], Y. Langevin[14], B. Sicardy[13], D. L. Matson[4], V. Formisano[6], P. Drossart[13] & V. Mennella[15]

Hyperion, Saturn's eighth largest icy satellite, is a body of irregular shape in a state of chaotic rotation[1,2]. The surface is segregated into two distinct units. A spatially dominant high-albedo unit having the strong signature of H_2O ice contrasts with a unit that is about a factor of four lower in albedo and is found mostly in the bottoms of cup-like craters. Here we report observations of Hyperion's surface in the ultraviolet and near-infrared spectral regions with two optical remote sensing instruments on the Cassini spacecraft at closest approach during a fly-by on 25–26 September 2005. The close fly-by afforded us the opportunity to obtain separate reflectance spectra of the high- and low-albedo surface components. The low-albedo material has spectral similarities and compositional signatures that link it with the surface of Phoebe and a hemisphere-wide superficial coating on Iapetus.

Cassini's Ultraviolet Imaging Spectrometer (UVIS) obtained the first ultraviolet maps of Hyperion, showing distinct variations in albedo across Hyperion's surface that are probably due to compositional differences (Fig. 1a). Particularly bright regions correspond with the rim of the large crater that dominated Cassini's views of Hyperion. The crater rims are bright, evidently as a result of exposure of relatively uncontaminated H_2O ice. Ultraviolet reflectance spectra from high-resolution data sets of Hyperion (Fig. 1b) show the differences in reflectance spectra of high- and low-albedo surface units. The UVIS reflectance spectra of Hyperion are dominated by the strong H_2O ice absorption feature that causes the precipitous change in reflectance at 165 nm (ref. 3) and which also appears in the low-albedo regions. In data from the Visible-Infrared Mapping Spectrometer (VIMS), the absorption bands of solid H_2O originally detected from ground-based observations[4] are confirmed, a band of complexed CO_2 is found, and other bands are also seen (data not shown).

Figure 2a shows the spectrum of the low-albedo material derived from the average of the 13 VIMS pixels from cube 1506393191 centred most closely on the darkest regions in the corresponding high-resolution ISS (Imaging Science Subsystem) image N1506393223. The 1.5- and 2.0-μm H_2O ice bands are weaker than in the bright areas (Fig. 2b), indicating a lower concentration of H_2O and/or larger grain sizes, and an additional absorption feature centred at 2.42 μm appears. This band was also seen in the spectra of Phoebe[5] and in the low-albedo material of Iapetus[6], where it was tentatively identified as an overtone band of a CN-bearing material, possibly KCN. The high-albedo material shows the strong spectral bands of solid H_2O in the ultraviolet (0.165 μm) and near-infrared (1.5, 2.0 and 3.0 μm); the
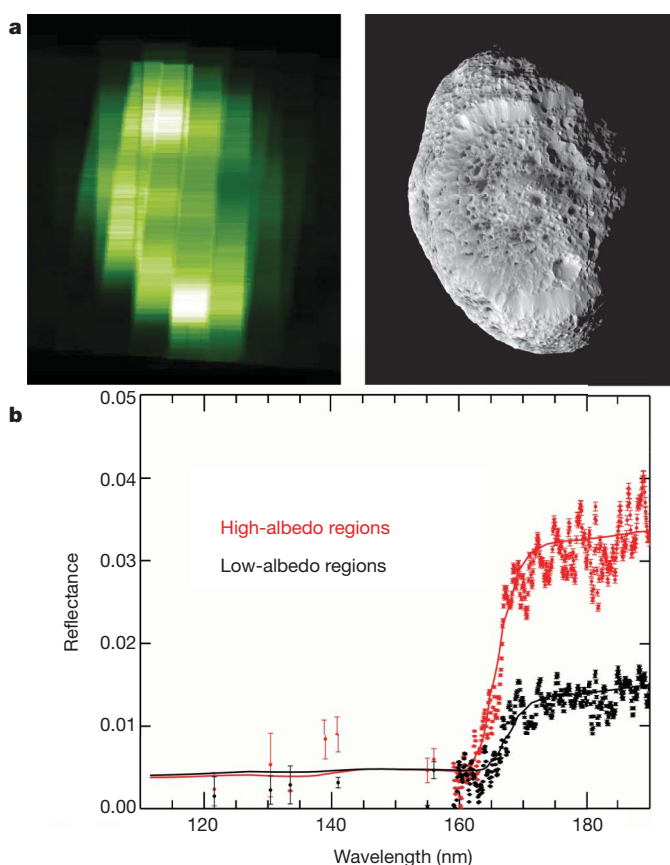

Figure 1 | Image and spectra from Cassini's UVIS. a, The UVIS image of Hyperion (left) compared with the visual appearance of the satellite at the same viewing geometry (right), obtained with the ISS cameras. In the ISS image, Hyperion is 300 km from top to bottom. The resolution in the UVIS images, taken in low-resolution mode, is ~55 km per pixel. The brightest pixels in the UVIS image are coincident with the broad expanses of high-albedo material on the large-scale, nearly circular scarp structure. **b**, Spectra of the two main albedo types of terrain on Hyperion, acquired during an observation when the resolution was ~6 km per pixel. Modelling details are given in Supplementary Information. Error bars, ±1σ.

[1]NASA Ames Research Center, MS 245-6, [2]USRA/SOFIA, NASA Ames Research Center, MS 211-3, Moffett Field, California 94035, USA. [3]SETI Institute, 515 N. Whisman Road, Mountain View, California 94043, USA. [4]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. [5]DLR, Institute for Planetary Exploration, Rutherfordstrasse 2, D-12489 Berlin, Germany. [6]INAF-IASF Istituto di Astrofisica Spaziale e Fisica Cosmica, [7]INAF-IASF Istituto dello Spazio Interplanetario, Via del Fosso del Cavaliere, 100, 00133 Roma, Italy. [8]USGS, Mail Stop 964, Box 25046, Denver Federal Center, Denver, Colorado 80225, USA. [9]Lunar and Planetary Laboratory and Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA. [10]Space Science Institute NW, 22 Fiddler's Road, Winthrop, Washington 98862-0667, USA. [11]Cornell University, 418 Space Sciences Building, Ithaca, New York 14853, USA. [12]University of Nantes, BP 92208, 2 rue de la Houssinière, 44072 Nantes Cedex 3, France. [13]Observatoire de Paris-Meudon, Département Recherche Spatial, 5 Place Jules Jannsen, 95129 Meudon Cedex, France. [14]Université de Paris Sud-Orsay, Institut d'Astrophysique Spatial, Batiment 120, 91405 Orsay Cedex, France. [15]INAF-OAC Osservatorio Astronomico di Capodimonte, Salita Moiariello 16, I-80131 Napoli, Italy.

band shapes and positions are consistent with crystalline H_2O, although a component of amorphous H_2O ice could also be present.

The VIMS data show the C–O stretching-mode band of the CO_2 molecule, although the band is shifted in wavelength from the nominal 4.27 μm to 4.25 μm, indicating that CO_2 on Hyperion is chemically or physically complexed with some other material (Supplementary Information). The CO_2 absorption band is evident in the low-albedo material (Fig. 2a), but is much weaker in the high-albedo ices (Fig. 2b). The central wavelength of this band is 4.25 μm, and the rounded shape of the band is different from the CO_2 bands in VIMS spectra of the low-albedo material on Iapetus[6] and Phoebe[5]. On all three satellites, the CO_2 band centre wavelength falls on the short-wavelength side of the 4.27-μm position observed in laboratory studies of pure CO_2 (ref. 7). In the spectrum of Hyperion's highest-albedo regions, the CO_2 band appears to be a fusion of the component seen in the low-albedo region spectrum (centred at 4.25 μm) and a component absorbing at 4.27 μm for pure CO_2 ice or frost. The spectra exhibiting the deepest CO_2 absorption features (approaching band depths of 45%) are not spatially correlated with craters or geologic features, suggesting that the ice is heterogeneous at these spatial scales.

The shifts in wavelength of the CO_2 bands seen on the saturnian satellites and on Ganymede and Callisto[8–10] strongly suggest that this molecule is complexed in one or more ways with another surface material. Neither pure CO_2 nor $CO_2 \cdot H_2O$ clathrate is stable for long times on the galilean or saturnian satellites because of their high vapour pressures at the temperatures of these bodies ($\sim$140 K and $\sim$90 K, respectively[11]), but other complexes may be.

Absorption bands attributed to aliphatic and aromatic hydrocarbons on Saturn's satellites Phoebe[5], Iapetus[6,12] and Enceladus[13] have been reported. The C–H stretching mode in polycyclic aromatic hydrocarbons (PAH) at 3.28 μm, detected in the low-albedo material of Iapetus[12], is of special interest because molecules of this type are found in interstellar dust, comets, meteorites and interplanetary dust particles. When aromatic molecules are embedded in H_2O ice and irradiated with ultraviolet light, new molecules of biological interest are created[14]. We have tentatively identified the PAH band on

Hyperion, noting that it is very weak and is distributed non-uniformly over the satellite's surface, as in the case of the other saturnian satellites on which organic spectral signatures have been identified. Turning to the 2.42-μm band, we note that it is most prominent in and around the low-albedo material. A number of nitriles (R–C≡N molecules) have overtone bands near this wavelength, but the best match thus far is with a band of KCN (ref. 5). Phoebe also has absorption features at 4.50 μm and longer wavelengths, which may be nitrile fundamentals. The strongest evidence for nitriles is the 2.42-μm band, although the specific molecule has not been identified.

We have modelled the spectrum of the average high- and low-albedo materials over the ultraviolet and near-infrared regions observed with UVIS and VIMS. In the ultraviolet, both the Hyperion high- and low-albedo regions can be fitted using intimate-mixture models with varying amounts of water ice and Triton tholin[15]. The components of the models of the near-infrared data are H_2O ice at temperature $T = 100$ K (in various grain sizes), and both Triton and Titan[16] tholins. The presence of this kind of complex, nitrogen-rich macromolecular organic solid material on the surface of Hyperion is consistent with its detection on other satellites of Saturn, notably Iapetus and Phoebe. The shapes of the spectra (reflectance levels, and red colour from 0.35 to 1.4 μm) in Figs 1 and 2 demonstrate that the organic material is distributed over both the high- and low-albedo surface units of Hyperion.

We have mapped the distribution over part of Hyperion's surface of H_2O ice (the 2.0-μm band; Fig. 3a), the material that absorbs at 2.42 μm (Fig. 3b), and the complex CO_2 (4.25 μm) absorption (Fig. 3c), by calculating the depths of the relevant absorption bands, assigning them colours (Fig. 3d), and then overlaying the resulting colour maps on the higher-resolution images from ISS (Fig. 3e, f). As seen in Fig. 3a and e, the H_2O ice bands are distinctly weaker in the low-albedo material than in the surrounding material of higher albedo. The strengths of the 2.42-μm band and the CO_2 band are correlated, producing yellow colour in the low-albedo material. CO_2 is not exclusively confined to the low-albedo material, appearing in association with H_2O ice throughout the observed areas (magenta colour). By contrast, the 2.42-μm feature is clearly strongest in and around the depressions, and apparent at lower levels in the icy surface immediately adjacent to the dark material as well.
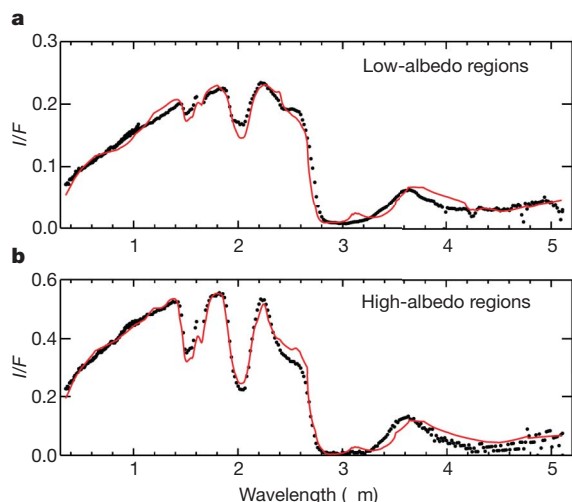

Figure 2 | Spectra and models from Cassini's VIMS. **a**, **b**, Spectral models calculated with scattering theory are compared with reflectance spectra of the low-albedo (**a**) and high-albedo (**b**) regions of Hyperion, extracted from VIMS observations at the closest approach to the satellite. The models include H_2O ice and two different examples of complex organic macromolecular carbonaceous material (tholins), providing a good overall match to the reflectance levels, spectral shapes, and H_2O band strengths seen in the data. See Supplementary Information for further details. The quantity I/F is the specific intensity of the surface; I is the intensity of the scattered sunlight, and πF is the incident solar flux normal (inclination = 0°) to the surface.
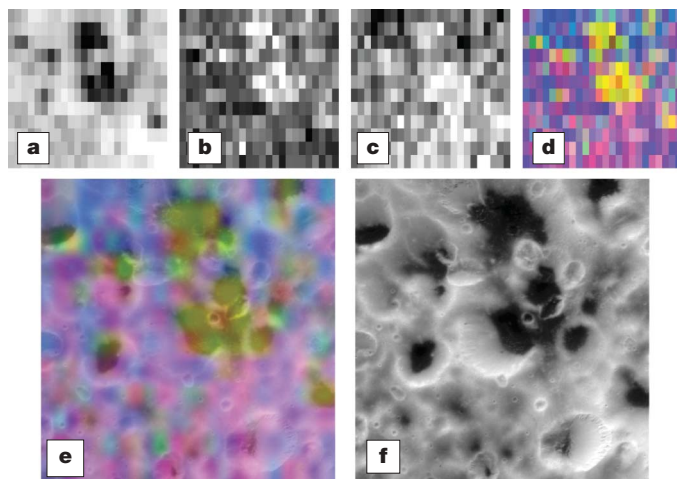
Figure 3 | Compositional map of a region of Hyperion. A colour-coded version of VIMS cube 1506393191 was made with three band-depth maps: the distribution of H_2O ice ('blue' (B), **a**), the 2.42-μm band attributed to CN ('green' (G), **b**), and the CO_2 band ('red' (R), **c**). In all cases, the deeper the band, the brighter the pixels. Panel **d** shows the RGB overlay of the three colours. Yellow indicates $CO_2 + CN$, with little H_2O; magenta indicates regions of $H_2O + CO_2$; and cyan denotes $CN + H_2O$. Panel **e** is the superposition of the VIMS colour map interpolated with a bi-cubic spline to the same spatial resolution (90 m per pixel) and the high-resolution ISS image (**f**; N1506393223) of the same field. The field of view is $\sim$38 km square.

Analysis of the ultraviolet and near-infrared spectra of the high- and low-albedo regions of Hyperion leads to several conclusions. The fundamental component of Hyperion's crust is a mix of solid H_2O and CO_2; this mixture is the country rock of the satellite. The H_2O exposed on the visible surface is primarily in a crystalline state, whereas the CO_2 is in some way complexed with other material (including, but not limited to, H_2O). The low-albedo materials on the floors of the cup-like craters exhibit enhanced CO_2 absorption, and a band attributed to an unidentified cyanide, with substantially reduced H_2O abundance. Maps of the distribution of the three principal absorbing materials also show the presence of the CO_2 and cyanide in the areas surrounding the cup-like craters.

The present investigation finds spectral similarities in the low-albedo material on both Hyperion and Iapetus, including complexed CO_2, the 2.42-μm band, widespread complex organic solid material—and possible aromatics on Hyperion. These results support the contention that Hyperion's colour and reflectance can be matched with a linear combination of H_2O ice and the material of low albedo and red colour found on the leading hemisphere of Iapetus[17,18]. Although the ultimate origin of the low-albedo material remains unsolved, the presence of organic material, H_2O ice, and complexed CO_2 implies compositional unity among these bodies, despite their significant differences in surface morphology and dynamical states.

1. Wisdom, J., Peale, S. J. & Mignard, F. The chaotic rotation of Hyperion. *Icarus* **58**, 137–152 (1984).
2. Thomas, P. & Veverka, J. Hyperion — Analysis of Voyager observations. *Icarus* **65**, 414–424 (1985).
3. Hendrix, A. R. & Hansen, C. J. Ultraviolet observations of Phoebe from Cassini UVIS. *Icarus* (in the press)..
4. Cruikshank, D. P. & Brown, R. H. Surface composition and radius of Hyperion. *Icarus* **50**, 82–87 (1982).
5. Clark, R. N. *et al.* Compositional mapping of Saturn's moon Phoebe with imaging spectroscopy. *Nature* **435**, 66–69 (2005).
6. Buratti, B. J. *et al.* Cassini visual and infrared mapping spectrometer observations of Iapetus: Detection of CO_2. *Astrophys. J.* **622**, L149–L152 (2005).
7. Sandford, S. A. & Allamandola, L. J. The physical and infrared spectral properties of CO_2 in astrophysical ice analogs. *Astrophys. J.* **355**, 357–372 (1990).
8. McCord, T. B. *et al.* Non water-ice constituents in the surface material of the icy Galilean satellites from the Galileo near-infrared mapping spectrometer investigation. *J. Geophys. Res.* **103**, 8603–8626 (1998).
9. Hibbitts, C. A. & Szanyi, J. Physisorption of CO_2 on non-ice materials of relevance to icy satellites. *Lunar. Planet. Sci. Conf.* **XXXVII**, 1753 (2006).
10. Chaban, G. M., Bernstein, M. & Cruikshank, D. P. Carbon dioxide on planetary bodies: Theoretical and experimental studies of molecular complexes. *Icarus* **187**, 592–599 (2007).
11. Lebofsky, L. A. Stability of frosts in the solar system. *Icarus* **25**, 205–217 (1975).
12. Cruikshank, D. P. *et al.* Hydrocarbons on Saturn's satellites Iapetus and Phoebe. *Icarus* (in the press)..
13. Brown, R. H. *et al.* Composition and physical properties of Enceladus' surface. *Science* **311**, 1425–1428 (2006).
14. Bernstein, M. P. *et al.* UV irradiation of polycyclic aromatic hydrocarbons in ices: Production of alcohols, quinones, and ethers. *Science* **283**, 1135–1138 (1999).
15. McDonald, G. D., Thompson, W. R., Heinrich, M., Khare, B. N. & Sagan, C. Chemical investigation of Titan and Trion tholins. *Icarus* **108**, 137–145 (1994).
16. Khare, B. N. *et al.* Optical constants of organic tholins produced in a simulated Titanian atmosphere — From soft X-ray to microwave frequencies. *Icarus* **60**, 127–137 (1984).
17. Jarvis, K. S., Vilas, F., Larson, S. M. & Gaffey, M. J. Are Hyperion and Phoebe linked to Iapetus? *Icarus* **146**, 125–132 (2000).
18. Buratti, B. J., Hicks, M. D., Tryka, K. A., Sittig, M. S. & Newburn, R. L. High-resolution 0.33–0.92 μm spectra of Iapetus, Hyperion, Phoebe, Rhea, Dione, and D-type asteroids: How are they related? *Icarus* **155**, 375–381 (2002).

nature

LETTERS

Low-energy acoustic plasmons at metal surfaces

Bogdan Diaconescu[1], Karsten Pohl[1], Luca Vattuone[2], Letizia Savio[2], Philip Hofmann[3], Vyacheslav M. Silkin[4], Jose M. Pitarke[5], Eugene V. Chulkov[4], Pedro M. Echenique[4], Daniel Farías[6] & Mario Rocca[7]

Nearly two-dimensional (2D) metallic systems formed in charge inversion layers[1] and artificial layered materials[2,3] permit the existence of low-energy collective excitations[4,5], called 2D plasmons, which are not found in a three-dimensional (3D) metal. These excitations have caused considerable interest because their low energy allows them to participate in many dynamical processes involving electrons and phonons[3], and because they might mediate the formation of Cooper pairs in high-transition-temperature superconductors[6]. Metals often support electronic states that are confined to the surface, forming a nearly 2D electron-density layer. However, it was argued that these systems could not support low-energy collective excitations because they would be screened out by the underlying bulk electrons[7]. Rather, metallic surfaces should support only conventional surface plasmons[8]—higher-energy modes that depend only on the electron density. Surface plasmons have important applications in microscopy[9,10] and sub-wavelength optics[11–13], but have no relevance to the low-energy dynamics. Here we show that, in contrast to expectations, a low-energy collective excitation mode can be found on bare metal surfaces. The mode has an acoustic (linear) dispersion, different to the $q_{\parallel}^{1/2}$ dependence of a 2D plasmon, and was observed on Be(0001) using angle-resolved electron energy loss spectroscopy. First-principles calculations show that it is caused by the coexistence of a partially occupied quasi-2D surface-state band with the underlying 3D bulk electron continuum and also that the non-local character of the dielectric function prevents it from being screened out by the 3D states. The acoustic plasmon reported here has a very general character and should be present on many metal surfaces. Furthermore, its acoustic dispersion allows the confinement of light on small surface areas and in a broad frequency range, which is relevant for nano-optics and photonics applications.

We performed the experiment in an ultrahigh-vacuum apparatus at a base pressure of about 2×10^{-10} mbar, equipped with an angle-resolved high-resolution electron energy loss (EEL) spectrometer[14]. In most measurements the energy resolution was set to about 16 meV. The single-crystal Be sample was cut and mechanically polished along the (0001) plane. It was cleaned through repeated 0.5–1 keV Ne$^+$ sputtering cycles with the sample at 450 °C followed by annealing periods at 500 °C until the amount of oxygen on the surface was below the sensitivity threshold of Auger electron spectroscopy and a fairly sharp low-energy electron-diffraction pattern was obtained. At this stage EEL spectra still showed the presence of oxygen on the sample, characterized by losses at 80 and 120 meV. Further cleaning resulted in a reduction of the oxygen loss intensity until the threshold of about 0.3% of the elastic peak in specular geometry was reached, at which point no further improvement was possible. The trace amounts of oxygen detected, estimated

to be around 2% of a monolayer, did not increase significantly after hours of measurements (see Supplementary Information C for details).

All EEL experiments were performed at room temperature (295 K). Figure 1 shows typical angle-resolved EEL spectra taken along the $\bar{\Gamma}-\bar{M}$ direction (Fig. 2b) for positive values of the momentum transfer parallel to the surface, $q_{\parallel}$ (Fig. 2c). A broad peak is observed to disperse as a function of $q_{\parallel}$ with another non-dispersing loss peak

$E_i = 7.26$ eV; $\theta_s = 63.3°$

— $E_{loss} = 0.30$ eV; $q_{\parallel} = 0.06$ Å^{-1}
— $E_{loss} = 0.40$ eV; $q_{\parallel} = 0.08$ Å^{-1}
— $E_{loss} = 0.59$ eV; $q_{\parallel} = 0.11$ Å^{-1}
— $E_{loss} = 0.73$ eV; $q_{\parallel} = 0.12$ Å^{-1}
— $E_{loss} = 0.76$ eV; $q_{\parallel} = 0.14$ Å^{-1}
— $E_{loss} = 0.80$ eV; $q_{\parallel} = 0.16$ Å^{-1}
— $E_{loss} = 0.97$ eV; $q_{\parallel} = 0.19$ Å^{-1}
— $E_{loss} = 1.42$ eV; $q_{\parallel} = 0.25$ Å^{-1}
— $E_{loss} = 1.59$ eV; $q_{\parallel} = 0.28$ Å^{-1}

$E_i = 10.74$ eV; $\theta_s = 59.2°$

— $E_{loss} = 0.70$ eV; $q_{\parallel} = 0.12$ Å^{-1}
— $E_{loss} = 0.90$ eV; $q_{\parallel} = 0.18$ Å^{-1}
— $E_{loss} = 1.30$ eV; $q_{\parallel} = 0.23$ Å^{-1}
— $E_{loss} = 1.46$ eV; $q_{\parallel} = 0.25$ Å^{-1}
— $E_{loss} = 1.60$ eV; $q_{\parallel} = 0.29$ Å^{-1}
— $E_{loss} = 1.83$ eV; $q_{\parallel} = 0.33$ Å^{-1}
— $E_{loss} = 1.87$ eV; $q_{\parallel} = 0.35$ Å^{-1}
— $E_{loss} = 1.92$ eV; $q_{\parallel} = 0.37$ Å^{-1}
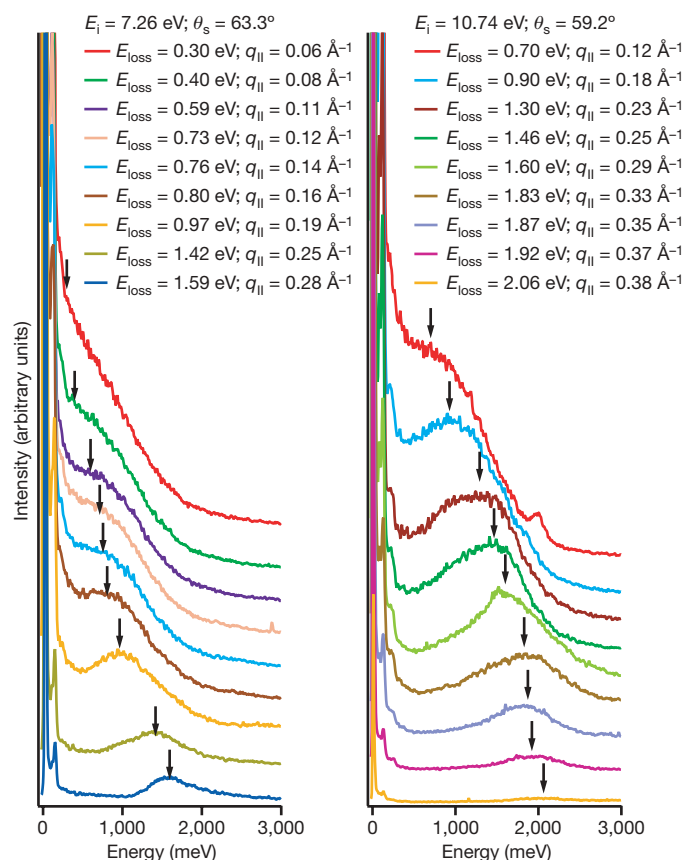— $E_{loss} = 2.06$ eV; $q_{\parallel} = 0.38$ Å^{-1}

Figure 1 | Families of angle-resolved EEL spectra. Spectra were taken at room temperature in the $\bar{\Gamma}-\bar{M}$ direction for two electron incident energies E_i and emergent scattering angles θ_s. The instrument employed a fixed analyser angle θ_s with a variable incident electron beam angle θ_i (ref. 14). Each spectrum corresponds to a different electron momentum-transfer component parallel to the surface $q_{\parallel}$. The spectra have been evenly spaced vertically for clarity. The additional, non-dispersing, low-energy loss is due to the residual oxygen contamination.

[1]Department of Physics and Material Science Program, University of New Hampshire, Durham, New Hampshire 03824, USA. [2]CNISM and Dipartimento di Fisica, Università di Genova, 16146 Genova, Italy. [3]Institute for Storage Ring Facilities and Interdisciplinary Nanoscience Center (iNANO), University of Aarhus, 8000 Aarhus C, Denmark. [4]Donostia International Physics Center (DIPC), Departamento de Fisica de Materiales and Centro Mixto CSIC-UPV/EHU, Facultad de Ciencias Quimicas, UPV/EHU, 20018 San Sebastian, Spain. [5]CIC nanoGUNE Consolider and Materia Kondentsatuaren Fisika Saila, UPV/EHU, Mikeletegi Pasealekua 56, E-2009 Donostia, Basque Country, Spain. [6]Departamento de Física de la Materia Condensada, Universidad Autónoma de Madrid, 28049 Madrid, Spain. [7]IMEM-CNR and Dipartimento di Fisica, Università di Genova, 16146 Genova, Italy.

due to traces of oxygen contaminants. The energy loss of the dispersing peak was determined via a multi-peak fitting procedure. The corresponding $q_{\parallel}$ was then calculated from energy and momentum conservation (see Supplementary Information B for a complete description of the method). This experimentally determined dispersion of the energy loss peak was measured up to 2 eV (Fig. 2a), well below the conventional Be surface plasmon energy of about 13 eV (refs 15, 16).

The experimental dispersion of the energy loss is clearly unaffected by changes in the scattering geometry and/or in incident energy of the electron beam. In the long-wavelength limit, we found the energy of the new mode to approach zero linearly for vanishing values of the momentum component parallel to the surface. Our data clearly show the acoustic character of this excitation within the limits of the experimental errors. We probed the surface at low $q_{\parallel}$ values of the first surface Brillouin zone (Fig. 2a and b). Owing to the isotropic surface state dispersion around $\bar{\Gamma}$ (Fig. 2d), the orientation of the electron scattering plane is not expected to influence the dispersion of the new excitation, as is confirmed by *ab initio* calculations for the $\bar{\Gamma} - \bar{K}$ direction. We tried to probe the new acoustic excitation for positive and negative $q_{\parallel}$. All the data shown are for positive momentum transfer. In the negative momentum transfer spectra, we saw no well-defined energy loss, presumably because the narrow dipole lobe results in a low excitation probability[17].

Metal surfaces such as Be(0001) and the (111) surfaces of noble metals support a partially occupied band of Shockley surface states with energies near the Fermi level. Their wavefunctions are strongly localized near the surface and decay exponentially into the solid, thus forming a quasi-2D electron gas overlapping the 3D bulk electrons. The use of a local dielectric function to describe the 3D continuum[7] indicates that a complete screening of a 2D charge density overlapping a 3D plasma may prevent the existence of any low-energy collective excitations. Here we show that if a full non-local dynamical screening at the surface due to underlying 3D bulk electrons is

considered, these experimental data can be interpreted as a novel collective electronic excitation (acoustic surface plasmon) of the quasi-2D surface charge distribution. This collective mode corresponds to out-of-phase charge-density oscillations of the 2D and 3D electron subsystems at a metal surface.

Information on collective electronic excitations at surfaces is obtained from the peak position of the imaginary part of the surface response function $g\left(\mathbf{q}_{\parallel}, \omega\right)$, which depends on the two-dimensional momentum transfer parallel to the surface $\mathbf{q}_{\parallel}$ and the frequency ω (refs 18, 19) (we use atomic units so that $e^2 = \hbar = m_e = 1$):

$$g\left(\mathbf{q}_{\parallel}, \omega\right) = \int d\mathbf{r} \int d\mathbf{r}' e^{q_{\parallel} z} \chi(\mathbf{r}, \mathbf{r}', \omega) V_{ext}(\mathbf{r}', \omega) \qquad (1)$$

where the external potential is of the form $V_{ext}(\mathbf{r}', \omega) = -\frac{2\pi}{q_{\parallel}} e^{q_{\parallel} z'} e^{i\mathbf{q}_{\parallel} \cdot \mathbf{r}'_{\parallel}} e^{-i\omega t}$ and the non-local frequency-dependent density-response function of the interacting electron system $\chi(\mathbf{r}, \mathbf{r}', \omega)$ is calculated in the framework of time-dependent density-functional theory using the integral equation (in symbolic form) $\chi = \chi^0 + \chi^0(v + f_{xc})\chi$, where v is the bare Coulomb potential, χ^0 represents the density-response function of non-interacting electrons, and f_{xc} is the exchange-correlation kernel chosen here to be zero (random-phase approximation). We calculate first the single-particle energies and wavefunctions that describe the surface band structure. With these wavefunctions and energies we then compute χ^0 and then solve the integral equation for χ (see Supplementary Information A).

The black dashed line shown in Fig. 2a is the predicted acoustic dispersion curve, assuming a free-electron like behaviour for the surface state on Be(0001) located in a wide 3D energy gap around the $\bar{\Gamma}$ point (ref. 16). The calculation agrees qualitatively with the experiment in the sense that both have an acoustic character, but the quantitative agreement is rather poor, owing to insufficient accuracy in describing the surface state dispersion. We are able to reproduce the experimental dispersion quantitatively by using an *ab initio* description of the surface electronic structure and the
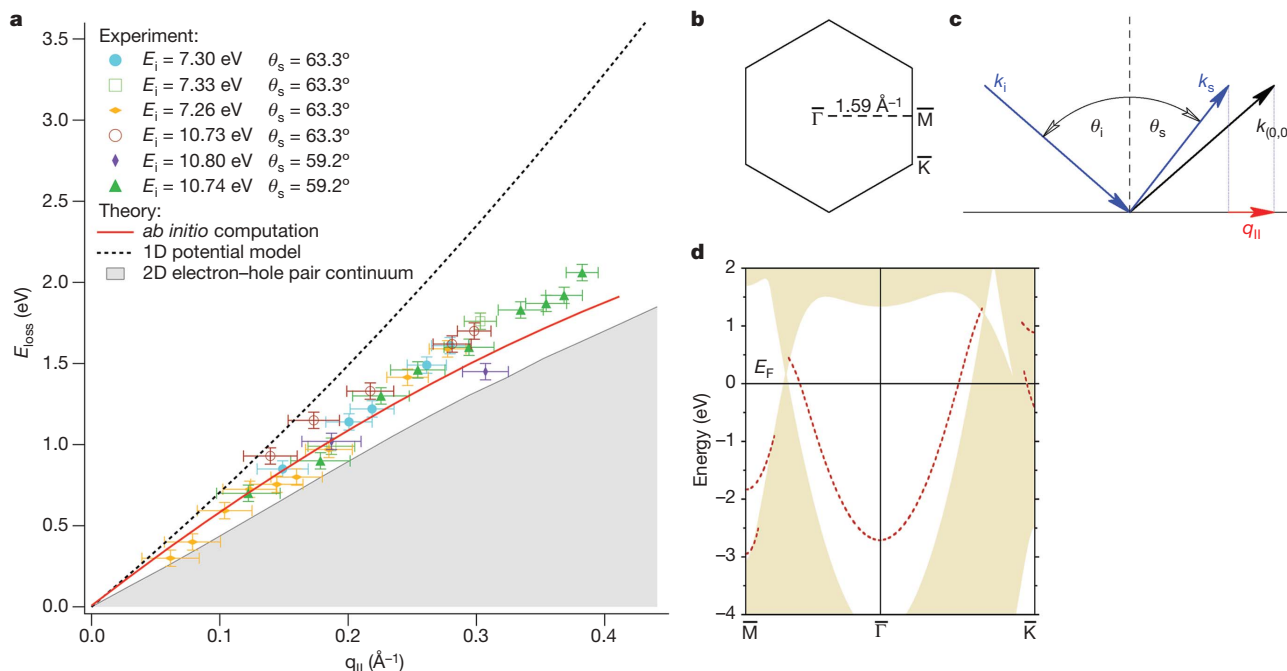

Figure 2 | Acoustic surface plasmon energy dispersion. a, The experimental dispersion was measured at room temperature and various incident electron energies and scattering angles. Energy error bars are due to uncertainties in the multi-peak deconvolution procedure of the EEL spectra, while $q_{\parallel}$ error bars represent the momentum integration window due to the finite angular acceptance of the EEL spectrometer (as described in Supplementary Information). The theoretical dispersion is indicated by the

black dashed line, showing the predicted acoustic surface plasmon dispersion obtained for a free-electron-like surface state, and by the solid red line, which was calculated using an *ab initio* Be(0001) surface band structure. **b,** First surface Brillouin zone of Be(0001). **c,** Scattering geometry in EEL spectra measurements. **d,** *Ab initio* Be(0001) surface electronic band structure. The red dotted line is the *ab initio* calculated surface state and the shaded area corresponds to the projected bulk states.

surface response function. The proper surface state dispersion around $\bar{\Gamma}$ (Fig. 2d) deviates from the free-electron scenario. In the occupied part it is nearly parabolic with a binding energy of 2.7 eV at $\bar{\Gamma}$, in close agreement with photoemission measurements and previous calculations[20–22]. Nevertheless, two important differences between the actual surface-state band and a band of free electrons are (1) the considerable deviation from parabolic behaviour above the Fermi level and (2) the abrupt cut at the borders of the energy gap around 1 eV above the Fermi level. Using the *ab initio* surface state dispersion as a starting point for a calculation of the acoustic surface plasmon dispersion results in the red line in Fig. 2a. The agreement with the experimental data is much better, greatly increasing our confidence in the interpretation.

The acoustic surface plasmon results from the interplay of the partially occupied electronic surface state (acting as a 2D electron density overlapping in the same region of space with the bulk electron gas) and the long-range Coulomb interaction manifested in the form of 3D dynamical screening of the 2D surface electron density. It corresponds to the out-of-phase charge oscillations between 2D and 3D subsystems and its dispersion is determined mainly by the surface-state Fermi velocity v_F^{2D} and closely follows the upper edge of the continuum for electron–hole pair excitations within the surface-state band (Fig. 2a). The Be(0001) surface, which has a high v_F^{2D}, is favourable for an experimental observation of the acoustic surface plasmon because the new collective excitation is well-defined up to relatively high energies of more than 1 eV. On other surfaces with a partially occupied surface-state band, such as the noble-metal (111) surfaces, the new mode is expected to be best defined at lower energies up to about several hundred millielectronvolts[23], thus making its EEL spectra detection more difficult[24]. The scattering geometry and incident energy need to be carefully chosen such that the scanned energy loss will cross the acoustic dispersion curve at low enough loss energies to prevent the occurrence of electron–hole transitions from occupied 3D bulk states to unoccupied 2D surface states. For electron–hole dynamics, however, this restriction is of small relevance because the low-energy region is much more important.

Acoustic surface plasmons owe their existence to the non-local screening of surface electrons caused by bulk electrons, at surfaces characterized by a partially occupied surface-state band lying in a wide bulk energy gap (Fig. 2d), and so acoustic surface plasmons should be a common phenomenon on many metal surfaces. Moreover, the acoustic plasmon dispersion closely follows the upper edge of electron–hole excitation, so it will affect the electron dynamics near the Fermi level much more than will conventional 2D plasmons[2], which, owing to their $q_{\parallel}^{1/2}$ dispersion, overlap in a much narrower range in energy–momentum space with the electron–hole continuum. Exciting a collective mode at very low energies can therefore lead to new situations at metal surfaces, arising from the competition between the incoherent electron–hole excitations and the new collective coherent mode. Many phenomena, such as electron, phonon and adsorbate dynamics as well as chemical reactions with characteristic energies lower than a few electronvolts, could be significantly influenced by the opening of a new low-energy decay channel such as the acoustic surface plasmon.

Of particular interest is the interaction of the acoustic surface plasmon with light. The slope of the acoustic surface plasmon dispersion, determined by v_F^{2D}, is about three orders of magnitude lower than the speed of light, and therefore the direct excitation of the new collective mode by light is not possible. However, nanometre-size objects at surfaces, such as atomic steps or molecular structures, can provide coupling between acoustic surface plasmons and light. The acoustic dispersion allows, at the same photon energy, for a collective surface excitation with a much lower associated wavelength than a conventional 2D plasmon with its $q_{\parallel}^{1/2}$ dispersion. In this way, the new mode can serve as a tool to confine light in a broad frequency range up to optical frequencies on surface areas of a few nanometres, thus ensuring the control of events at metal surfaces with both high spatial (nanometres) and high temporal (femtoseconds) resolution. Another consequence of the acoustic character of the dispersion is that both phase and group velocity of the collective excitation are the same, so signals can be transmitted undistorted along the surface. Given that a theoretically estimated decay length of the ASP of 100–1,000 nm is expected for the medium (100 meV) to the far (10 meV) infrared, this is an appealing prospect for the field of nano-optics.

1. Allen, S. J., Tsui, D. C. & Logan, R. A. Observation of the two-dimensional plasmon in silicon inversion layers. *Phys. Rev. Lett.* **38**, 980–983 (1977).
2. Nagao, T., Hildebrandt, T., Henzler, M. & Hasegawa, S. Dispersion and damping of a two-dimensional plasmon in a metallic surface-state band. *Phys. Rev. Lett.* **86**, 5747–5750 (2001).
3. March, N. H. & Tosi, M. P. Collective effects in condensed conducting phase including low-dimensional systems. *Adv. Phys.* **44**, 299–386 (1995).
4. Stern, F. Polarizability of a two-dimensional electron gas. *Phys. Rev. Lett.* **18**, 546–548 (1967).
5. Chaplik, A. V. Possible crystallization of charge carriers in low-density inversion layers. *Sov. Phys. JETP* **35**, 395–398 (1972).
6. Ruvalds, J. Are plasmons the key to superconducting oxides? *Nature* **328**, 299 (1987).
7. Sarma, S. D. & Madhukar, A. Collective modes of spatially separated, two-component, two-dimensional plasma in solids. *Phys. Rev. B* **23**, 805–815 (1981).
8. Ritchie, R. H. Plasma losses by fast electrons in thin films. *Phys. Rev.* **106**, 874–881 (1957).
9. Schuster, S. C., Swanson, R. V., Alex, L. A., Bourret, R. B. & Simon, M. I. Assembly and function of a quaternary signal transduction complex monitored by surface plasmon resonance. *Nature* **365**, 343–347 (1993).
10. Flatgen, G. *et al.* Two-dimensional imaging of potential waves in electrochemical systems by surface plasmon microscopy. *Science* **269**, 668–671 (1995).
11. Barnes, W. L., Dereux, A. & Ebbesen, T. W. Surface plasmon subwavelength optics. *Nature* **424**, 824–830 (2003).
12. Lezec, H. *et al.* Beaming light from a subwavelength aperture. *Science* **297**, 820–822 (2002).
13. Pendry, J. Playing tricks with light. *Science* **285**, 1687–1688 (1999).
14. Rocca, M., Valbusa, U., Gussoni, A., Maloberti, G. & Racca, L. Apparatus for adsorption studies. *Rev. Sci. Instrum.* **62**, 2172–2176 (1991).
15. Höchst, H., Steiner, P. & Hüfner, S. The conduction electron hole coupling in beryllium metal. *Phys. Lett.* **60A**, 69–71 (1977).
16. Silkin, V. M. *et al.* Novel low-energy collective excitation at metal surfaces. *Europhys. Lett.* **66**, 260–264 (2004).
17. Rocca, M. Low-energy EELS investigation of surface electronic excitations on metals. *Surf. Sci. Rep.* **22**, 1–71 (1995).
18. Persson, B. N. J. & Zaremba, E. Electron-hole pair production at metal surfaces. *Phys. Rev. B* **31**, 1863–1872 (1985).
19. Liebsch, A. *Electronic Excitations at Metal Surfaces* (Plenum, London, 1997).
20. Karlsson, U. O., Flodström, S. A., Engelhardt, R., Gädeke, W. & Koch, E. E. Intrinsic surface state on Be(0001). *Solid State Commun.* **49**, 711–714 (1984).
21. Bartynski, R. A., Jensen, E., Gustafsson, T. & Plummer, E. W. Angle-resolved photoemission investigation of the electronic structure of Be: Surface states. *Phys. Rev. B* **32**, 1921–1926 (1985).
22. Chulkov, E. V., Silkin, V. M. & Shirykalov, E. N. Surface electronic structure of Be(0001) and Mg(0001). *Surf. Sci.* **188**, 287–300 (1987).
23. Silkin, V. M., Pitarke, J. M., Chulkov, E. V. & Echenique, P. M. Acoustic surface plasmons in the noble metals Cu, Ag, and Au. *Phys. Rev. B* **72**, 115435–115441 (2005).
24. Politano, A., Chiarello, G., Formoso, V., Agostino, R. & Colavita, E. Plasmon of Shockley surface states in Cu(111): A high-resolution electron energy loss spectroscopy study. *Phys. Rev. B* **74**, 081401(R) (2006).

LETTERS

Early geochemical environment of Mars as determined from thermodynamics of phyllosilicates

Vincent Chevrier[1], Francois Poulet[2] & Jean-Pierre Bibring[2]

Images of geomorphological features that seem to have been produced by the action of liquid water have been considered evidence for wet surface conditions on early Mars[1]. Moreover, the recent identification of large deposits of phyllosilicates, associated with the ancient Noachian terrains[2,3] suggests long-timescale weathering[4] of the primary basaltic crust by liquid water[2,5]. It has been proposed that a greenhouse effect resulting from a carbon-dioxide-rich atmosphere sustained the temperate climate required to maintain liquid water on the martian surface during the Noachian[6,7]. The apparent absence of carbonates and the low escape rates of carbon dioxide[8], however, are indicative of an early martian atmosphere with low levels of carbon dioxide. Here we investigate the geochemical conditions prevailing on the surface of Mars during the Noachian period using calculations of the aqueous equilibria of phyllosilicates. Our results show that Fe^{3+}-rich phyllosilicates probably precipitated under weakly acidic to alkaline pH, an environment different from that of the following period, which was dominated by strongly acid weathering[9] that led to the sulphate deposits identified on Mars[10–12]. Thermodynamic calculations demonstrate that the oxidation state of the martian surface was already high, supporting early escape of hydrogen. Finally, equilibrium with carbonates implies that phyllosilicate precipitation occurs preferentially at a very low partial pressure of carbon dioxide. We suggest that the possible absence of Noachian carbonates more probably resulted from low levels of atmospheric carbon dioxide, rather than primary acidic conditions[13]. Other greenhouse gases may therefore have played a part in sustaining a warm and wet climate on the early Mars.

Phyllosilicates, especially clay minerals, are usual products of the weathering process, resulting from the interaction between the atmosphere, the hydrosphere and the lithosphere. Impact-driven hydrothermalism may also form clays in the subsurface[14]. Although this process is unlikely to explain the OMEGA (Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité) observations of deposits of several hundred square kilometres not correlated to impacts[2], it still involves liquid water, mostly from meteoric or surface origin. Here we discuss a potential aqueous weathering process that might have leached the primary minerals to form the phyllosilicates detected on the surface of Mars. During leaching, fractionation of elements occurs because water removes ions progressively from primary silicates, according to their solubility. Soluble ions like K^+, Na^+, Ca^{2+} or Mg^{2+} are dissolved first, leaving less soluble ions like Fe^{3+}, Al^{3+} and H_4SiO_4 (silica) to precipitate as phyllosilicates[15]. Therefore, phyllosilicates can be used as indicators of the degree of ion leaching by liquid water on the surface of Mars. The detection by OMEGA of Mg/Fe smectites primarily, as well as montmorillonite, $Ca_{0.167}(Al_{1.67}Mg_{0.33})Si_4O_{10}(OH)_2$ (ref. 2), containing both insoluble (Fe^{3+}, Al^{3+}) and soluble (Mg^{2+} and Ca^{2+}) ions, indicates weak fractionation of elements and thus moderate leaching. In neutral to alkaline conditions (see below), the calculated solubility of smectites is less than $10^{-7}\,g\,l^{-1}$. The combination of low solubility with moderate leaching suggests that the weathering process occurred at a moderate water-to-rock ratio. Very high water-to-rock ratios would have led to Al-rich phyllosilicates such as kaolinite (only tentatively detected once by OMEGA). Indeed, kaolinite is deprived of soluble ions and contains only Al^{3+} and silica, indicating a drastic leaching of all soluble ions, and its very low solubility ($10^{-14}\,g\,l^{-1}$ at pH = 7) allows it to precipitate in extremely dissolved systems.

The main chemical parameters most likely to control the precipitation of phyllosilicates are mineral, water and atmospheric compositions, as well as the temperature of the solution. The formation of iron smectite nontronite mostly depends on the oxidation–reduction and pH conditions (Fig. 1). In addition to these fundamental parameters, aluminium, iron and silica activities have a strong influence on the precipitation of nontronite. Berthierine is the usual Fe^{2+}-phyllosilicate in reducing pedogenetic environments[4,16,17]. The equilibrium between berthierine and nontronite is mainly dependent on the aluminium concentration. The presence of nontronite is only possible when the activity of aluminium becomes sufficiently low, at most 10^{-5} (Fig. 1a). Thus, at high Fe (0.8×10^{-3}) and low Al^{3+} (10^{-25}) activities, nontronite precipitates at pH values between 4 and 10, that is, in slightly acidic to alkaline conditions. Nontronite also involves precipitation of the Fe^{3+} phase rather than the Fe^{2+} phase, indicating much lower iron solubility (activity of 10^{-10}). In such higher oxidation conditions, nontronite is forced to precipitate at higher pH values of at least 6 (Fig. 1b). However, nontronite could also precipitate in subsurface environments[2], in equilibrium with endogenic Fe^{2+}-phyllosilicates such as greenalite or minnesotaite (Fig. 1c). In this case, the most significant difference is the appearance of ferrihydrite in the system, and the possibility that nontronite could precipitate at even higher pH values, up to 12 (Fig. 1c). Such environments usually present higher temperatures, which shift the stability field of nontronite towards a lower pH (about one unit at 100 °C, Fig. 1c).

Silica activity has a strong influence on smectite precipitation. High silica activities allow nontronite to precipitate only at low pH, down to 3 at the saturation relative to amorphous silica (Fig. 2). Such low-pH conditions are often associated with the formation of sulphates observed at the surface of Mars[9,11,12]. Indeed, the presence of SO_2 generates strongly acidic waters that inhibit the formation of nontronite, and favours Fe^{3+} sulphates. Partial pressures of SO_2 as low as 10^{-8} bar generate waters with pH values of ~2–3 and high SO_4^{2-} activities. These values are clearly in the stability field of jarosite (Fig. 3)[18]. In conclusion, the presence of nontronite indicates the following conditions: (1) a low aluminium activity, (2) a high oxidation level, (3) weakly acidic to alkaline conditions and (4) very low abundance of SO_2.

[1]W. M. Keck Laboratory for Space Simulation, Arkansas Center for Space and Planetary Sciences, MUSE 202, University of Arkansas, Fayetteville, Arkansas 72701, USA. [2]Institut d'Astrophysique Spatiale, Université Paris-Sud and CNRS (UMR 8617), F-91405, Orsay, France.

The highly oxidizing conditions required for the precipitation of nontronite imply oxidation of the Fe^{2+} in primary silicates into Fe^{3+}. O_2 and H_2O_2 are the best potential oxidant. However, O_2 abundance in the early martian atmosphere probably remained very low, below 10^{-12} bar (ref. 19). On Earth, it took about 2 billion years for the O_2 level to reach today's high values[20], mostly because of biological activity and photosynthesis. Even in the present oxidizing martian atmosphere, the H_2O_2 density is extremely low[21] (6×10^{-15} cm^{-2}). The most abundant potential oxidant on early Mars is water. Oxidation of Fe^{2+} into Fe^{3+}, even if not thermodynamically

Figure 1 | pE–pH diagrams of nontronite stability at 25 °C with other phases most likely to be present on Mars. These phases are: aqueous Fe^{2+} and Fe^{3+}, berthierine, minnesotaite, greenalite $Fe^{2+}_3Si_2O_5(OH)_4$ and ferrihydrite. **a**, Berthierine, with the following aqueous species activities: $log[Fe] = -3.1$, $log[H_4SiO_4] = -4.5$, $log[Al^{3+}] = -25$, $log[Ca^{2+}] = -3.3$. The black dashed line shows the stability fields for the upper limit of aluminium activity (10^{-5}) needed to still have nontronite. If the activity of Al^{3+} is larger than 10^{-5}, berthierine precipitates in almost all conditions, except for very low pH (below 3) and the stability field of nontronite is extremely restricted (indicated by black arrow). **b**, Low activitiy of iron (10^{-10}) resulting from oxidative conditions, indicated by the presence of Fe^{3+} phases, and for the following silica activities: $log[H_4SiO_4] = -4.5$ (black lines), $log[H_4SiO_4] = -5.5$ (dashed lines). **c**, Nontronite forms in equilibrium with endogenic phyllosilicates (minnesotaite and greenalite (dashed lines)) and for $log[Fe] = -10$, $log[H_4SiO_4] = -4.5$, $log[Al^{3+}] = -25$. Endogenic conditions can lead to higher temperatures, so the equilibrium diagram is also calculated with minnesotaite at 100 °C (dotted lines) in the same conditions, using the Van't Hoff relationship[30], assuming that the entropy variation is negligible in the temperature interval[30].
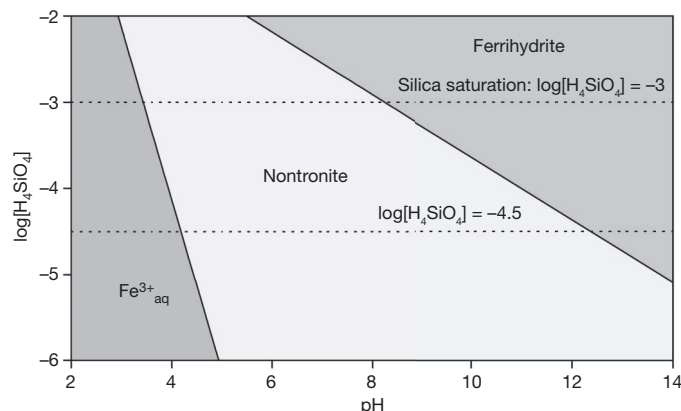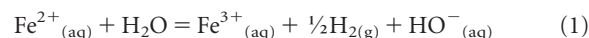

Figure 2 | Stability field of nontronite, aqueous Fe^{3+} and ferrihydrite as a function of dissolved silica activity and pH. Activities of Ca^{2+} and Al^{3+} are the same as for Fig. 1a. The maximum of silica activity is 10^{-3}, where amorphous silica starts to precipitate, whereas 3.2×10^{-5} ($log[H_4SiO_4] = -4.5$) corresponds to the medium silica activity used to build the pE–pH diagrams showed in Fig. 1.

favoured, may be driven by thermal escape of H_2, resulting from the reduction of water[22,23], according to the following equation:

$$Fe^{2+}_{(aq)} + H_2O = Fe^{3+}_{(aq)} + \tfrac{1}{2}H_{2(g)} + HO^-_{(aq)} \quad (1)$$

A greenhouse effect of CO_2 can be invoked to maintain liquid water long enough to allow silicate alteration, and phyllosilicate formation. Such a primitive environment has often been proposed to explain the multiple geomorphological evidences of fluvial activity[6] and may have resulted from primary degassing[7,24]. Abundant CO_2 should have led to formation of carbonate deposits, which have not yet been detected by OMEGA. It has been proposed that early Noachian acidic conditions prevented precipitation of carbonates[13]. However, according to our calculations, the smectites observed by OMEGA on the Noachian crust indicate zones where weathering of primary silicates has acted in slightly acidic to alkaline conditions, much more favourable to carbonate formation. The large phyllosilicate-rich deposits are spatially disconnected from the Hesperian
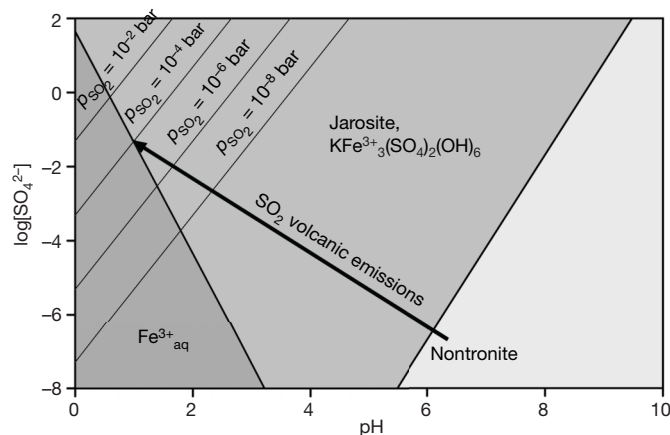

Figure 3 | Equilibrium of nontronite versus jarosite as a function of dissolved sulphate activity (SO_4^{2-}) and pH. Activities of Fe, Al^{3+}, H_4SiO_4 and Ca^{2+} are the same as for Fig. 1a, and $log[K^+] = -5$. The thin lines represent various partial pressures of SO_2 in equilibrium with water, at a H_2 partial pressure of 10^{-6} bar, according to the following reaction: $SO_2 + H_2O = SO_4^{2-} + 2H^+ + H_2$. The thick arrow represents the evolution of the conditions with increasing pressure of SO_2 if the dissolution of SO_2 is stoichiometric, forcing surface conditions to evolve from neutral-alkaline (nontronite) to acidic and sulphate-rich conditions (jarosite). The large volcanic activity at the end of the Noachian[5,26] has probably inhibited the precipitation of smectites, inducing the formation of the sulphate deposits widely observed in the Hesperian terrains.

sulphate-rich areas[3], where the corresponding pH conditions were too low at ~3–4 (refs 9 and 25) to allow smectite precipitation (Fig. 1).

The late Noachian to early Hesperian (around 3.7 Gyr ago) intense volcanism resulting from the Tharsis rise[26] and other volcanoes has probably injected large amounts of SO_2 into the atmosphere, ending the neutral-weathering period of smectite formation, possibly dissolving the carbonates if present, and shifting the conditions into the sulphate-acidic environment responsible for the sulphates observed in later Hesperian terrains (Fig. 3). Smectites would still be present after this acidic phase during the Hesperian because they are far more resistant to redissolution than carbonates, and mechanical erosion continuously exposes fresh outcrops. Our modelling can also explain *in situ* observations of the Fe^{3+}-bearing phase made by the Mars Exploration Rovers Spirit and Opportunity[27,28] as well as the detection of small phyllosilicate-rich outcrops by OMEGA in Terra Meridiani[2]. These observations indicate localized (in space and time) environments in which dissolved silica activity might have been sufficiently high to produce both phyllosilicates and sulphates. This could result from evaporation processes, efficient in increasing the concentrations of dissolved species[9].

The (pH–p_{CO_2}) aqueous equilibrium between carbonates and smectites provides a direct insight into the CO_2 partial pressure during their formation. The abundance of Mg/Fe-rich smectites[2] probably reflects the primary mineralogy of the martian crust dominated by Mg and Fe silicates. Because the precipitation of nontronite depends on the oxidation–reduction conditions and the pH, the equilibrium between carbonates and Mg,Fe-smectites is a function of the Fe abundance (expressed by the Mg number $x_{Mg} = Mg/(Mg+Fe)$, in mol), and p_{CO_2} (Fig. 4). If only Mg-smectite saponite, $Ca_{0.165}Mg_3(Al_{0.33}Si_{3.67})O_{10}(OH)_2$, precipitates, the equilibrium is

independent of the pH value, at a p_{CO_2} of 2.5×10^{-3} bar. Increasing Fe (decreasing x_{Mg}) abundances towards the nontronite end member makes the equilibrium more sensitive to the pH. In this case the Fe-rich smectite can precipitate at higher p_{CO_2} if the pH is higher, and at lower p_{CO_2} in a less neutral environment (Fig. 4). It has been previously demonstrated that the precipitation of nontronite is also strongly sensitive to the silica activity (Fig. 2). Setting the silica activity at the saturation relative to amorphous silica (10^{-3}) provides an upper limit for the p_{CO_2}. Lower silica activity logically favours carbonates, which sets the carbonate–smectite boundary at lower p_{CO_2} (Fig. 4).

The ALH84001 meteorite is an Mg,Fe-orthopyroxenite ($x_{Mg} \approx 0.8$) of Noachian age. If it is representative of the primary martian crust composition, the p_{CO_2} of the equilibrium between carbonates and smectites is low between 10^{-2} and 10^{-3} atm at silica saturation, and 10^{-5} to 10^{-4} for the lower silica activity (3.2×10^{-5}, Fig. 4). Therefore, the presence of smectites without carbonates, if confirmed when analysed at sample levels (NASA's Mars Science Laboratory and the ESA's Exomars, which are next-generation rovers for the surface exploration of Mars), would suggest that the p_{CO_2} remained low during the Noachian, and thus probably during the entire history of Mars.

A low p_{CO_2} value in the Noachian is not sufficient to sustain a warm and wet atmosphere. Therefore, a sufficient greenhouse effect might require the presence of other gases, such as sulphur dioxide, ammonia and methane. However, ammonia is rapidly photolysed to N_2 and H_2 (ref. 29). The absence of sulphate deposits in the Noachian terrains[5] does not favour a sulphate-rich early Noachian atmosphere (Fig. 3), leaving methane as an alternative[17], provided that a large flux was present.

METHODS

The smectite compositions represent chemical end members most likely to be present on the early Mars[30], according to its primary mafic composition. The thermodynamic values used in the calculations come from various sources (see Supplementary Information). All diagrams were calculated by setting the equilibrium reactions and calculating the corresponding Gibbs free energy of the reaction $\Delta_r G^0_{298 K, 1 bar}$, converted into the reaction constant K (see equations in the Supplementary Information) according to the formula:

$$K = \exp[-\Delta_r G^0_{298 K, 1 bar}/RT] \qquad (2)$$

where R is the ideal gas constant and T is the temperature of the reaction. The pH of the equilibrium was calculated from the products of the activities of the dissolved species. For the oxidation–reduction reaction potential E, we used the following formula:

$$E = E^0 + (0.059/n)\log K \qquad (3)$$

where n is the number of exchanged electrons. The equilibrium reactions are then plotted in a space in which E is a function of pH such that voltage E divided by 0.059 gives a pH equivalent, that is, an activity of electrons in the solution (pE) rather than a voltage (Fig. 1). Eventually other variables can be used, such as activities of important ions (such as H_4SiO_4 in Fig. 2 or SO_4^{2-} in Fig. 3) or partial pressures (p_{CO_2} in Fig. 4). The temperature effect was determined using the Van't Hoff equation, assuming that the enthalpy variation $\Delta_r H^0_{298 K, 1 bar}$ is negligible in the temperature interval[30]:

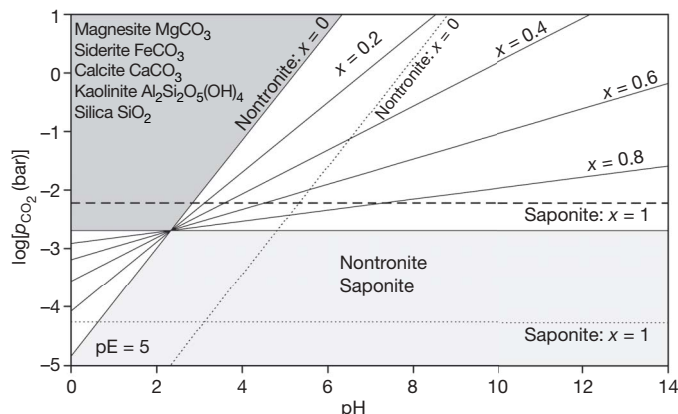$$\ln(K_2/K_1) = (\Delta_r H^0_{298 K, 1 bar}/R) \times (1/T_1 - 1/T_2) \qquad (4)$$

Figure 4 | Equilibrium diagram between carbonates and smectites as a function of the CO_2 partial pressure, the pH and the Mg number, x_{Mg}. ($x_{Mg} = 1$ corresponds to Mg-smectite saponite, whereas $x_{Mg} = 0$ corresponds to Fe-smectite nontronite). Calcite and kaolinite represent accessory phases associated to the Fe,Mg-carbonates (siderite and dolomite), accounting for the excess of Ca^{2+} and Al^{3+} present in the Mg,Fe-smectites. The oxidation potential pE has been set up at 5 (oxidizing water). The silica activity has been set up at the highest value (10^{-3}) corresponding to the saturation relative to amorphous silica, giving an upper limit to the CO_2 partial pressures (thin black lines). If we consider the silica activity of the initial model (3.2×10^{-5}), the equilibrium lines of smectites and carbonates (thin dotted lines) are shifted about two orders of magnitude towards lower CO_2 partial pressures. Both dark grey and light grey zones correspond to exclusive ranges of conditions (for high silica activity) where only carbonates or smectites, respectively, can exist. In the white zone, the equilibrium depends on the composition in Mg and Fe of the smectite/carbonate assemblage. Increasing concentration in Mg (and the Mg number) makes the equilibrium less sensitive to the pH. The horizontal dashed line indicates today's conditions on Mars. Considering the usually quite large error on thermodynamic values (about 0.5 log units), present-day conditions are very close to the equilibrium between Mg-rich smectites and carbonates.

1. Malin, M. C. & Edgett, K. S. Evidence for persistent flow and aqueous sedimentation on early Mars. *Science* **302**, 1931–1934 (2003).
2. Poulet, F. *et al.* Phyllosilicates on Mars and implications for the early Mars history. *Nature* **481**, 623–627 (2005).
3. Bibring, J.-P. *et al.* Mars surface diversity as revealed by the OMEGA/Mars Express observations. *Science* **307**, 1576–1581 (2005).
4. Meunier, A. & El Albani, A. The glauconite–Fe-illite–Fe-smectite problem: a critical review. *Terra Nova* **19**, 95–104 (2006).

5. Bibring, J. P. *et al.* Global mineralogical and aqueous Mars history derived from OMEGA/Mars express data. *Science* **312**, 400–404 (2006).

6. Jakosky, B. M. & Phillips, R. J. Mars' volatile and climate history. *Nature* **412**, 237–244 (2001).

7. Pollack, J. B., Kasting, J. F., Richardson, S. M. & Poliakoff, K. The case for a wet, warm climate on early Mars. *Icarus* **71**, 203–224 (1987).

8. Barabash, S., Fedorov, A., Lundin, R. & Sauvaud, J. A. Martian atmospheric erosion rates. *Science* **315**, 501–503 (2007).

9. Tosca, N. J. *et al.* Geochemical modeling of evaporation processes on Mars: insight from the sedimentary record at Meridiani Planum. *Earth Planet. Sci. Lett.* **240**, 122–148 (2005).

10. Arvidson, R. E. *et al.* Spectral reflectance and morphologic correlations in Eastern Terra Meridiani, Mars. *Science* **307**, 1591–1594 (2005).

11. Gendrin, A. *et al.* Sulfates in Martian layered terrains: the OMEGA/Mars Express view. *Science* **307**, 1587–1591 (2005).

12. Squyres, S. W. *et al.* In situ evidence for an ancient aqueous environment at Meridiani Planum, Mars. *Science* **306**, 1709–1714 (2004).

13. Fairen, A. G. *et al.* Inhibition of carbonate synthesis in acidic oceans on early Mars. *Nature* **431**, 423–426 (2004).

14. Naumov, M. V. Principal features of impact-generated hydrothermal circulation systems: mineralogical and geochemical evidence. *Geofluids* **5**, 165–184 (2005).

15. Eggleton, R. A., Foudoulis, C. & Varkevisser, D. Weathering of basalt: changes in rock chemistry and mineralogy. *Clays Clay Mineral.* **35**, 161–169 (1987).

16. Sheldon, N. D. & Retallack, G. J. Low oxygen levels in earliest Triassic soils. *Geology* **30**, 919–922 (2002).

17. Sheldon, N. D. Precambrian paleosols and atmospheric CO_2 levels. *Precambr. Res.* **147**, 148–155 (2006).

18. Elwood-Madden, M. E., Bodnar, R. J. & Rimstidt, J. D. Jarosite as an indicator for water-limited weathering on Mars. *Nature* **431**, 821–823 (2004).

19. Catling, D. C. & Moore, J. M. The nature of coarse-grained crystalline hematite and its implications for the early environment of Mars. *Icarus* **165**, 277–300 (2003).

20. Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).

21. Encrenaz, T. *et al.* Hydrogen peroxide on Mars: evidence for spatial and seasonal variations. *Icarus* **170**, 424–429 (2004).

22. Chevrier, V., Rochette, P., Mathé, P.-E. & Grauby, O. Weathering of iron rich phases in simulated Martian atmospheres. *Geology* **32**, 1033–1036 (2004).

23. Lammer, H. *et al.* Loss of water from Mars: implications for the oxidation of the soil. *Icarus* **165**, 9–25 (2003).

24. Phillips, R. J. *et al.* Ancient geodynamic and global-scale hydrology on Mars. *Science* **291**, 2587–2591 (2001).

26. Solomon, S. C. *et al.* New perspectives on ancient Mars. *Science* **307**, 1214–1220 (2005).

27. Wang, A. *et al.* Evidence of phyllosilicates in Wooly Patch, an altered rock encountered at West Spur, Columbia Hills, by the Spirit rover in Gusev crater, Mars. *J. Geophys. Res.* **111**, E02S16, doi:10.1029/2005JE002516 (2006).

28. Glotch, T. D. *et al.* Mineralogy of the light-toned outcrop at Meridiani Planum as seen by the Miniature Thermal Emission Spectrometer and implications for its formation. *J. Geophys. Res.* **111**, E12S03, doi:10.1029/2005JE002672 (2006).

29. Sagan, C. & Chyba, C. The early faint Sun paradox: organic shielding of ultraviolet-labile greenhouse gases. *Science* **276**, 1217–1221 (1997).

30. Gooding, J. L. Chemical weathering on Mars. Thermodynamic stabilities of primary minerals (and their alteration products) from mafic igneous rocks. *Icarus* **33**, 483–513 (1978).

LETTERS

Co-mimics have a mutualistic relationship despite unequal defences

Hannah M. Rowland[1], Eira Ihalainen[2], Leena Lindström[2], Johanna Mappes[2] & Michael P. Speed[1]

In the first clear mathematical treatment of natural selection, Müller[1] proposed that a shared warning signal (mimicry) would benefit defended prey species by sharing out the per capita mortality incurred during predator education. Although mimicry is a mainstay of adaptationist thinking, there has been repeated debate on whether there is a mutualistic or a parasitic relationship between unequally defended co-mimic species[2–5]. Here we show that the relationship between unequally defended species is mutualistic. We examined this in a 'novel world'[6] of artificial prey with wild predators (great tit, *Parus major*). We kept the abundance of a highly defended prey ('model') constant and increased the density of a moderately defended prey ('defended mimic') of either perfect or imperfect mimetic resemblance to the model. Both model and defended mimic showed a net benefit from a density-dependent decrease in their per capita mortality. Even when the effect of dilution through density was controlled for, defended mimics did not induce additional attacks on the model, but we found selection for accurate signal mimicry. In comparison, the addition of fully edible (batesian) mimics did increase additional attacks on the model, but as a result of dilution this resulted in no overall increase in per capita mortality. By ignoring the effects of density, current theories may have overestimated the parasitic costs imposed by less defended mimics on highly defended models.

Since Müller's original statement of his mimicry theory in 1878 (ref. 1), müllerian mimicry has been explored comparatively[7], theoretically[8] and empirically[9–14]. Yet there remain ambiguities at the heart of the theory, leading to continued controversy and debate[5,8,15–17]. At its core, Müller's theory assumes that, to learn avoidance of prey with a particular visual signal, predators need to kill a fixed number of that prey. Hence, two distasteful species that have separate signals pay a higher mortality cost during predator learning than two species that, through mimicry, resemble each other. In contrast, Bates[18] proposed that perfectly palatable mimics could exist because they gained protection by exploiting predators' learned avoidance of defended model species. So although müllerian mimicry is considered to be beneficial to all mimetic prey, in batesian mimicry the model–mimic relationship is typically considered parasitic, because edible mimics degrade the predators' association of the shared signal with defence[19,20]. Despite the superficial simplicity of these mimicry theories, considerable confusion still exists in the literature, particularly surrounding the dynamics of an intermediate case between batesian and müllerian mimicry (in which both prey are defended but to unequal extents)[2,4,15,21,22]. Some authors argue that moderately defended species may act parasitically and dilute the protection of the better defended species[4], whereas others argue that the relationship is always mutualistic[5,8]. Furthermore, some theoretical models predict that changes in the total densities of unequally defended mimetic prey can change their relationship from parasitism in low abundances into mutualism at higher abundances[3,17,23]

Despite the broad interest in these questions, there is no decisive experimental evidence that directly tests the opposing hypotheses. In this experiment we sought to resolve whether inequality in defence causes a mutualistic or a parasitic mimetic relationship when changes in prey densities are also incorporated. We tested the predation pressure imposed by predators that hunt visually (wild-caught great tits, *Parus major*) in a laboratory-based 'novel world'. In this environment the individual great tits were naive predators. We modified the densities of artificial mimetic prey that were unequally defended; a highly defended prey (here termed 'model') accompanied by a moderately defended prey (here termed 'defended mimic') or an edible mimic ('edible mimic'). In addition to this, we manipulated the degree of visual similarity of the mimic to the model (either 'perfect' or 'imperfect' mimicry, see Table 1). At the start of each trial, 60 edible cryptic prey were present as alternative prey (see Table 1). The highly defended model was either presented alone (120 prey at the start of each trial, model plus cryptic prey) or accompanied by 30 or 60 defended or edible mimics. Trials terminated when 50 prey had been 'killed' by a bird, simulating a predator with fixed-quota foraging. In addition, because mathematical simulations of mimicry often use time-based foraging[4], we analysed the data to simulate a time-limited forager (see Supplementary Information).

The birds responded differently to the highly and moderately defended prey items when these were presented alone, eating significantly more of the moderately defended mimic than the highly defended model (independent sample *t*-test, $t_{18} = -2.110$, $P = 0.049$). We found that increasing the number of defended mimics (treatments 1, 3, 4, 5 and 6) decreased the mortality of the models (Fig. 1a; general linear model (GLM), $F_{(2,44)} = 9.625$, $P < 0.001$), and this was true for both visually perfect and imperfect defended mimics (no effect of signal accuracy ($F_{(1,44)} = 0.283$, $P = 0.597$) or

Table 1 | Experimental setup

	Perfect defended mimics				Imperfect defended mimics			Edible mimics		
Signal	Type	Treatment			Signal	Type	Treatment	Signal	Type	Treatment
		1	2	3	4			5	6			7	8	9
⊠	Cryptic	60	60	60	60	⊠	Cryptic	60	60	⊠	Cryptic	60	60	60
■	Model	60	–	60	60	■	Model	60	60	■	Model	–	60	60
■	Mimic	–	60	30	60	◆	Mimic	30	60	■	Mimic	60	30	60

The types and numbers of prey presented in the experimental treatments are shown. An alternative edible prey was cryptic (a cross) as it matched the crosses on the aviary background where prey was presented (see Supplementary Fig. S1). Models always had a black square signal and were highly unpalatable. Mimics were either squares (perfect mimicry in treatments 3 and 4 and 8 and 9) or diamonds (imperfect mimicry in treatments 5 and 6), and were either moderately defended (treatments 2–6) or edible (treatments 8 and 9). Numbers in the columns correspond to the number of each prey type presented at the start of a trial.

[1]School of Biological Sciences, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK. [2]Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35, FI-40014 Jyväskylä University, Finland.

interaction between density and signal accuracy ($F_{(2,44)} = 0.026$, $P = 0.974$)). Hence the model benefited similarly from the presence of defended mimics regardless of whether the mimic's signal was perfect or imperfect (see also ref. 24).

The perfect defended mimics also had a decreased per capita mortality as their status as mimics changed from being non-mimetic at a density of 60 prey (presented without model; treatment 2) to being mimetic at a density of 30 or 60 mimics (with model; treatments 3 and 4; Fig. 1b; general linear model (GLM), $F_{(2,27)} = 11.432$, $P < 0.001$). Mimics did not show significantly lower mortality as they increased abundance from 30 to 60 prey (treatments 3–6; see Fig. 1b; general linear model (GLM), $F_{(1,36)} = 3.509$, $P = 0.069$). Imperfect defended mimics showed higher mortality than the perfect mimics (Fig. 1b, blue bars; $F_{(1,36)} = 5.843$, $P = 0.021$), indicating selection for perfect mimicry in the moderately defended mimics. There was no two-way interaction between mimic density and signal perfection ($F_{(1,36)} = 0.000$, $P = 1.000$; Fig. 1b).

Because the total density of prey increased when defended mimics were added to the setup, the probability of attack for a member of the population of models (the likelihood of being eaten by chance) decreased. To control for dilution we calculated the risk of attack for the models relative to that predicted assuming random predation (see Methods). This shows whether the models suffered increased attacks but to an extent that would have been masked in the measure of per capita mortality as a result of increased total density of prey. We found that the addition of defended mimics did not affect the mean risk of predation for models (no effect of mimic density (general linear model (GLM), $F_{(2,44)} = 1.478$, $P = 0.239$), signal accuracy ($F_{(1,44)} = 0.256$, $P = 0.615$) or interaction between density and signal accuracy ($F_{(2,44)} = 0.022$, $P = 0.979$)). However, when we examined the learning process closely (see Supplementary Information) we found that the risk of predation for the model was lower at the end of the trial (Fig. 1c; last ten prey items killed) when paired with perfect compared with imperfect defended mimics (general linear model (GLM), $F_{(1,36)} = 5.861$, $P = 0.021$; no effect of density ($F_{(1,36)} = 0.487$, $P = 0.490$) and no interaction between density and signal accuracy ($F_{(1,36)} = 0.325$, $P = 0.572$); see also ref. 13 and Supplementary Information for a detailed analysis of the learning process). This shows that, in addition to dilution, accuracy of mimicry is a source of benefit for the models. In terms of relative mortality, the benefit of signal similarity was greater for the defended mimics (see Fig. 1b).

Because the effect of the difference in defence levels fits within the original müllerian framework (that is, the defended mimics benefit rather than harm the models), we tested whether our results match the quantitative predictions of Müller's original model (which assumes equality in defence between co-mimics). Müller assumed that if n individuals of mimetic species 1 and 2 are killed during predator learning, the number of each species taken is proportionate to their abundances (a_1 and a_2, respectively). We estimated two values of n: one was based on highly defended models alone (treatment 1) and the other on moderately defended mimics alone (treatment 2; see Supplementary Information). From these values of n the numbers of models and defended mimics killed (a_1 and a_2) could be predicted for the treatments in which both co-mimics were present. This allowed us to test whether the birds behaved as though all mimetic prey was of one defence level. We found that when mimics were added, the birds attacked more models than would be predicted if all prey in the model–mimic mixture were responded to as though they were highly defended (n estimated from models in Supplementary Table S1). The numbers killed were a better match to the prediction where n was based on the moderately defended mimics (see Supplementary Table S2). This can be explained by the mild taste of the mimics, which may have degraded learning (but see ref. 24), or by the tendency of predators to kill more aposematic prey (in absolute numbers) when prey are more numerous, even if all prey are highly defended[25]. The number of perfect mimics killed was intermediate, being not significantly different from either set of predictions (n based on models or defended mimics, Supplementary Tables S3 and S4); however, imperfect defended mimics (treatments 5 and 6 in Supplementary Tables S3 and S4) were taken at a higher rate than would be predicted if n were based on models. Despite some disproportionate costs incurred during learning, the result remains that mortalities decreased through dilution.

We next asked whether the strong dilution effect could extend to fully edible (batesian) mimetic prey. Abundant batesian mimics are generally considered to have a negative effect on their model's survival (see, for example, refs 18, 20, 26, 27). However, it has been argued that edible mimics could prove beneficial to their models if they increase in abundance to the point of satiating the predator[17,23]. To test these opposing predictions, we increased the number of edible mimics in the setup while keeping the number of models constant (overall density increased with the addition of edible mimics).

We found that increasing the number of edible mimics had no effect on the mortality of models (Fig. 2a; GLM univariate ANOVA, $F_{(1,27)} = 0.015$, $P = 0.904$) but affected the mortality of edible mimics (Fig. 2b; GLM univariate ANOVA, $F_{(1,25)} = 44.619$, $P = 0.001$). The edible mimics had lower mortality when associated with models
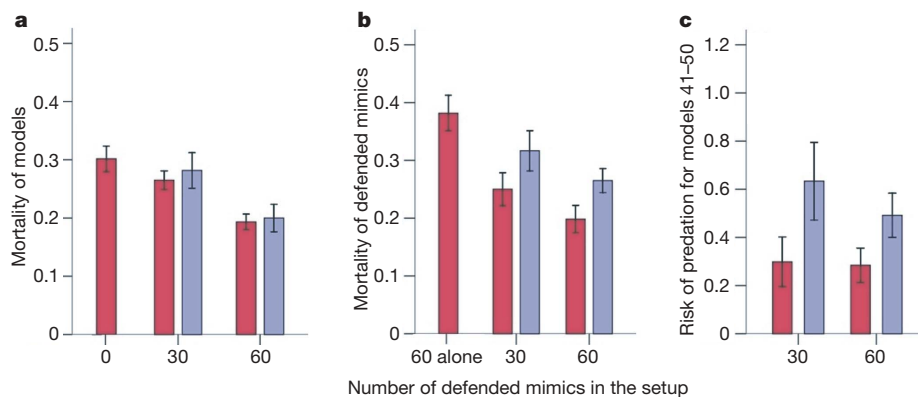

Figure 1 | Per capita mortality of models and defended mimics, and relative predation risk of models. a, Mortality of models when presented alone with 60 cryptic edible prey (mimics in setup 0) and when presented with 30 or 60 visually perfect (squares, see Table 1) defended mimics (red columns) or visually imperfect defended mimics (blue columns; diamonds, see Table 1). **b,** Mortality of visually perfect defended mimics (red columns) and visually imperfect defended mimics (blue columns) when presented alone with cryptic edible prey, or at a density of 30 or 60 with 60 models. **c,** The relative risk of predation for models in the last ten prey killed when paired with perfect (red columns) and imperfect (blue columns) defended mimics. All results are shown as means ± s.e.m.; $n = 10$ in all treatments.

(Tukey post-hoc tests; 60 mimics alone versus 60 models plus 30 mimics (treatments 7 and 8), $P < 0.001$, and 60 mimics alone versus 60 models plus 60 mimics (treatments 7 and 9), $P < 0.001$). However, the mortality of edible mimics was not affected by increases in their density from 30 to 60 ($P = 0.429$).

It does not necessarily follow that batesian mimics have no detrimental effects on their models. The absolute numbers of signalling prey killed (models and edible mimics together) increased with increasing numbers of edible mimics (Fig. 2c; GLM univariate ANOVA, $F_{(3,27)} = 5.67$, $P = 0.004$). To examine how relative predator choices changed with the addition of batesian mimics, we again calculated the mean relative risk of predation (see Methods). The relative risk for models increased with increasing number of edible mimics (Fig. 2d; GLM univariate ANOVA, $F_{(2,27)} = 3.57$, $P = 0.042$). This result also shows that the presence of edible mimics was detrimental to the models. Hence, in terms of relative risk of predation there is some evidence of a cost of batesian mimicry to the model, but the net effect of the edible mimics on per capita mortality is no change.

Thus, we found that the net effect of mimicry between unequally defended müllerian co-mimics was not parasitic. However, batesian mimics degraded the model's protection but this effect was more than compensated for through density-dependent dilution, which could be a real ecological benefit. As illustrated by the comparison of per capita mortality and relative predation risk, different measures can reveal different but equally relevant aspects of between-species dynamics (see also ref. 28). The comparison also raises the important question of predation pressure: if population sizes of mimetic prey change, how does the predator community react? Perhaps an increased density of defended prey does not attract additional predators as much as undefended prey (such as batesian mimics). The implication is that mimetic relationships between defended prey may be less sensitive than parasitic (batesian) relationships to changes at the community level (see also refs 29, 30). Our results clearly illustrate that the classic example of mimicry as an adaptation cannot be understood without explicitly considering the population and community context. Both the relative and absolute abundances of the species in the mimetic complex can have profound effects on mimetic dynamics in ways not clearly understood in the current mimicry literature.

METHODS SUMMARY

Ninety-seven great tits (*Parus major*) were pretrained to handle artificial prey items (pieces of almond in paper shells) and to forage in a 'novel world' aviary (see methods in ref. 24). There were nine treatments (Table 1); highly defended models were mimicked by, first, visually perfect, moderately defended mimics; second, visually imperfect, moderately defended mimics; or third, visually perfect, edible mimics (batesian mimics). Cryptic alternative prey was also presented (Table 1). High and moderate unpalatability of the prey items provided different levels of defence to the birds; they ate more of the moderately defended mimic than the highly defended model when these were presented alone with the cryptic prey (Table 1 and Fig. 1; see also refs 24, 28). The models and perfect mimics had a black square as a warning signal, whereas imperfect mimics had a diamond-shaped signal (Table 1). The numbers of models and cryptic prey were kept constant at 60. The models were presented either alone or accompanied by 30 or 60 mimics (moderately defended or edible; Table 1). Each bird was required to 'kill' (eat or taste) 50 prey items in the trial. We recorded the absolute numbers of all prey types killed and calculated the per capita mortality of each prey type by dividing the total number of each prey type killed during the trial by the number initially presented. We also calculated relative predation risk by dividing the number of each prey type taken by the predicted number that would have been killed assuming random predation (that is, in which prey types are eaten according to their frequencies). For example, because the predators were allowed to kill 50 prey items during the trial, the expected predation for models when presented with only cryptic prey was 25, whereas in treatment 3 the expected predation would be 20 models, 10 defended mimics and 20 edible crosses.
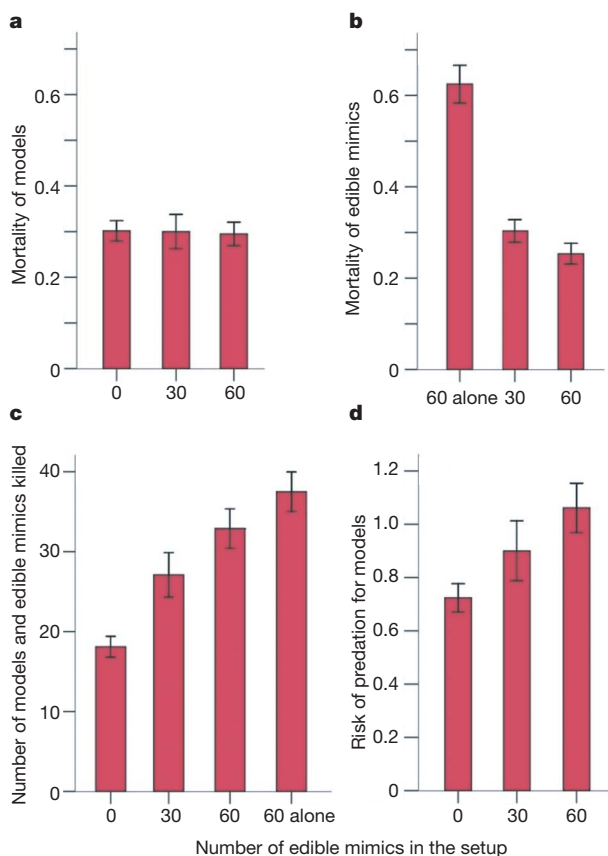
Figure 2 | Per capita mortality of models and edible mimics, total numbers of models and edible mimics killed, and relative predation risk for models. a, Mortality of models when presented alone with 60 cryptic edible prey (mimics in setup 0), and when presented with either 30 or 60 edible mimics. **b,** Mortality of edible mimics when presented alone with 60 cryptic edible prey (60 alone) or at a density of 30 or 60 with 60 models. **c,** The absolute number of models and edible mimics attacked when models were presented with 60 edible cryptic prey (mimics in setup 0), or when models were presented with either 30 or 60 edible mimics, and for comparison the treatment where edible mimics were presented alone with cryptic edible prey (60 alone, $n = 8$). **d,** The mean relative risk of predation for models when presented alone with 60 cryptic edible prey (mimics in setup 0) or with 30 or 60 edible mimics. All results are shown as means ± s.e.m.; $n = 10$ for all treatments, except where stated otherwise.

1. Müller, F. Ueber die Vortheile der Mimicry bei Schmetterlingen. *Zool. Anz.* **1,** 54–55 (1878).
2. Dixey, F. A. Presidential Address, Section D. *Rep. Br. Assoc. Advmt Sci.* 199–207 (1919).
3. Owen, R. E. & Owen, A. R. G. Mathematical paradigms for mimicry—recurrent sampling. *J. Theor. Biol.* **109,** 217–247 (1984).
4. Speed, M. P. Muellerian mimicry and the psychology of predation. *Anim. Behav.* **45,** 571–580 (1993).
5. Mallet, J. & Joron, M. Evolution of diversity in warning color and mimicry: Polymorphisms, shifting balance, and speciation. *Annu. Rev. Ecol. Syst.* **30,** 201–233 (1999).
6. Alatalo, R. V. & Mappes, J. Tracking the evolution of warning signals. *Nature* **382,** 708–710 (1996).
7. Dumbacher, J. P. & Fleischer, R. C. Phylogenetic evidence for colour pattern convergence in toxic pitohuis: Müllerian mimicry in birds? *Proc. R. Soc. Lond. B* **268,** 1971–1976 (2001).
8. Mallet, J. Causes and consequences of a lack of coevolution in Müllerian mimicry. *Evol. Ecol.* **13,** 777–806 (1999).
9. Kapan, D. D. Three-butterfly system provides a field test of müllerian mimicry. *Nature* **409,** 338–340 (2001).
10. Symula, R., Schulte, R. & Summers, K. Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Müllerian mimicry hypothesis. *Proc. R. Soc. Lond. B* **268,** 2415–2421 (2001).

11. Benson, W. W. Natural selection for Müllerian mimicry in *Heliconius erato* in Costa Rica. *Science* **176**, 936–939 (1972).
12. Mallet, J. & Barton, N. H. Strong natural selection in a warning-colour hybrid zone. *Evolution Int. J. Org. Evolution* **43**, 421–431 (1989).
13. Rowe, C., Lindström, L. & Lyytinen, A. The importance of pattern similarity between Müllerian mimics in predator avoidance learning. *Proc. R. Soc. Lond. B* **271**, 407–413 (2004).
14. Speed, M. P., Alderson, N. J., Hardman, C. & Ruxton, G. D. Testing Müllerian mimicry: an experiment with wild birds. *Proc. R. Soc. Lond. B* **267**, 725–731 (2000).
15. Sherratt, T. N., Speed, M. P. & Ruxton, G. D. Natural selection on unpalatable species imposed by state-dependent foraging behaviour. *J. Theor. Biol.* **228**, 217–226 (2004).
16. Kokko, H., Mappes, J. & Lindström, L. Alternative prey can change model–mimic dynamics between parasitism and mutualism. *Ecol. Lett.* **6**, 1068–1076 (2003).
17. Speed, M. P. Batesian, quasi-Batesian or Müllerian mimicry? Theory and data in mimicry research. *Evol. Ecol.* **13**, 755–776 (1999).
18. Bates, H. W. Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconidae. *Trans. Linn. Soc. Lond.* **23**, 495–566 (1862).
19. Sheppard, P. M. *Natural Selection and Heredity* (Hutchinson, London, 1975).
20. Fisher, R. A. *The Genetical Theory of Natural Selection* 2nd edn (Dover Publications, New York, 1958).
21. Marshall, G. A. K. On diaposematism, with reference to some limitations of the Müllerian hypothesis of mimicry. *Trans. Entomol. Soc.* **1908**, 93–142 (1908).
22. Huheey, J. E. Studies in warning coloration and mimicry. VIII. Further evidence for a frequency-dependent model of predation. *J. Herpetol.* **14**, 223–230 (1980).
23. Sherratt, T. N. State-dependent risk-taking by predators in systems with defended prey. *Oikos* **103**, 93–100 (2003).
24. Ihalainen, E., Lindström, L. & Mappes, J. Investigating Müllerian mimicry: predator learning and variation in prey defences. *J. Evol. Biol.* **20**, 780–791 (2007).
25. Lindström, L., Alatalo, R. V., Lyytinen, A. & Mappes, J. Strong antiapostatic selection against novel rare aposematic prey. *Proc. Natl Acad. Sci. USA* **98**, 9181–9184 (2001).
26. Huheey, J. E. Mathematical models of mimicry. *Am. Nat.* **31**, s22–s41 (1988).
27. Sheppard, P. M. The evolution of mimicry: a problem in ecology and genetics. *Cold Spring Harb. Symp. Quant. Biol.* **24**, 131–140 (1959).
28. Lindström, L., Lyytinen, A., Mappes, J. & Ojala, K. Relative importance of taste and visual appearance for predator education in Müllerian mimicry. *Anim. Behav.* **72**, 323–333 (2006).
29. Nonacs, P. Foraging in a dynamic mimicry complex. *Am. Nat.* **126**, 165–180 (1985).
30. Lindström, L., Alatalo, R. V., Lyytinen, A. & Mappes, J. The effect of alternative prey on the dynamics of imperfect Batesian and Müllerian mimicries. *Evolution Int. J. Org. Evolution* **58**, 1294–1302 (2004).

Author Contributions All authors contributed equally to the planning of the work and writing the paper. H.M.R. and E.I. were the primary data collectors. First authorship was decided between H.M.R. and E.I.; authors after the first are listed in alphabetical order.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.M.R. (h.m.rowland@liv.ac.uk).

METHODS

Ninety-seven wild great tits (*Parus major*) were caught at feeding stations around Konnevesi Research Station and were used as visually hunting predators in the experiment, which was conducted from January to March 2005 (permission was granted by the Central Finland Regional Environmental Center and by the Experimental Animal Committee of the University of Jyväskylä). Each bird was familiarized with a large experimental aviary (57.7 $m^2 \times 3.5$ m height) and trained to open artificial prey items (see methods in ref. 24) and trained to forage from the aviary floor, which comprised a white background with black crosses printed onto it (see Supplementary Fig. S1). The artificial prey items used in the experiment (see Table 1) were made either highly or moderately unpalatable by soaking the almonds in a chloroquine phosphate solution for 1 h (2 g of chloroquine dissolved in 30 ml of water, or 0.25 g of chloroquine dissolved in 30 ml of water, respectively). These doses match those used in previous experiments[24,28]. The difference in concentration translated to a difference in palatability to great tits; the birds treated the highly and moderately defended prey items differently when they were presented alone, eating significantly more of the moderately defended mimic than the highly defended model (see the text and also refs 24, 28).

Learning experiment. We used three major conditions, split into nine experimental treatments, which comprised a highly defended model mimicked by, first, visually perfect, moderately defended mimics; second, visually imperfect, moderately defended mimics; or third, visually perfect, edible mimics (batesian mimics). The signal of the models and perfect mimics was a black square, whereas imperfect mimics had a diamond-shaped signal (Table 1). These signals do not differ in their visibility or efficacy as warning signals[24]. The number of models was kept constant at 60, and 60 edible cryptic prey items were also presented in all treatments as alternative prey. The highly defended model was either presented alone or accompanied by 30 or 60 mimics (see Table 1). Each bird was released into the experimental aviary, and the number and type of prey attacked were recorded. A prey item was noted as killed when the bird opened the paper shell and took a bite or ate the almond contained within the shell. Each bird was required to 'kill' 50 prey items before the trial was terminated.

Statistical analyses. In addition to the absolute numbers of prey killed, we calculated the per capita mortality of each prey type by dividing the total number of each type of prey killed during the trial by the number initially presented. We also calculated relative predation risk by dividing the number of each prey type taken by the predicted number that would be killed if predation were random, regardless of conspicuousness and defence levels. Predation is random when prey types are eaten according to their frequencies, taking into account the frequency of the cryptic prey. Because the models and mimics were equally visible, and the number of less conspicuous cryptic prey was kept constant, the visibility difference between the cryptic and signalling prey does not affect the comparison of risks for co-mimics between the different densities. The data did not require transformation to satisfy requirements of parametric statistics. The data were analysed by GLM with Tukey post-hoc pairwise comparisons in SPSS v.13.0. Details of further analyses are given in the Supplementary Information.

Nature

LETTERS

Mutation of *FIG4* causes neurodegeneration in the pale tremor mouse and patients with CMT4J

Clement Y. Chow[1], Yanling Zhang[2], James J. Dowling[4], Natsuko Jin[2], Maja Adamska[1], Kensuke Shiga[5], Kinga Szigeti[5,7], Michael E. Shy[9], Jun Li[9,10], Xuebao Zhang[9], James R. Lupski[5,6,8], Lois S. Weisman[2,3] & Miriam H. Meisler[1]

Membrane-bound phosphoinositides are signalling molecules that have a key role in vesicle trafficking in eukaryotic cells[1]. Proteins that bind specific phosphoinositides mediate interactions between membrane-bounded compartments whose identity is partially encoded by cytoplasmic phospholipid tags. Little is known about the localization and regulation of mammalian phosphatidylinositol-3,5-bisphosphate (PtdIns(3,5)P_2), a phospholipid present in small quantities that regulates membrane trafficking in the endosome–lysosome axis in yeast[2]. Here we describe a multi-organ disorder with neuronal degeneration in the central nervous system, peripheral neuronopathy and diluted pigmentation in the 'pale tremor' mouse. Positional cloning identified insertion of *ETn2β* (early transposon 2β)[3] into intron 18 of *Fig4* (*A530089I17Rik*), the homologue of a yeast SAC (suppressor of actin) domain PtdIns(3,5)P_2 5-phosphatase located in the vacuolar membrane. The abnormal concentration of PtdIns(3,5)P_2 in cultured fibroblasts from pale tremor mice demonstrates the conserved biochemical function of mammalian Fig4. The cytoplasm of fibroblasts from pale tremor mice is filled with large vacuoles that are immunoreactive for LAMP-2 (lysosomal-associated membrane protein 2), consistent with dysfunction of the late endosome–lysosome axis. Neonatal neurodegeneration in sensory and autonomic ganglia is followed by loss of neurons from layers four and five of the cortex, deep cerebellar nuclei and other localized brain regions. The sciatic nerve exhibits reduced numbers of large-diameter myelinated axons, slowed nerve conduction velocity and reduced amplitude of compound muscle action potentials. We identified pathogenic mutations of human *FIG4* (KIAA0274) on chromosome 6q21 in four unrelated patients with hereditary motor and sensory neuropathy. This novel form of autosomal recessive Charcot–Marie–Tooth disorder is designated CMT4J.

Mutant mice with severe tremor, abnormal gait and diluted pigmentation were identified in our mouse colony on a mixed inbred strain background. One breeding pair generated 8/30 affected progeny, consistent with inheritance of an autosomal recessive mutation designated 'pale tremor'. At postnatal day three (P3), affected homozygotes have diluted pigmentation and reduced size (Fig. 1a). Intentional tremor develops during the second week after birth, and abnormal limb postures are evident by the third week (Fig. 1b). The impaired motor coordination, muscle weakness and 'swimming' gait of pale tremor mice are demonstrated in the Supplementary Movie. There is progressive loss of mobility, reduction in body weight and juvenile lethality (Fig. 1c).

To genetically map the pale tremor gene, a cross was carried out with strain CAST/Ei. The recovery of affected F_2 offspring was 9% (50/532), indicative of prenatal loss on this genetic background. We genotyped 532 F_2 animals with microsatellite and single nucleotide polymorphism markers to map the pale tremor gene to a 2-megabase interval of mouse chromosome 10 between *D10Umi13* and *D10Mit184* (Supplementary Fig. 1a). There are 21 annotated genes in the non-recombinant interval (http://www.ensembl.org; mouse build 35); these were tested as candidates by sequencing reverse transcription polymerase chain reaction (RT–PCR) products from brain RNA. The *A530089I17Rik* transcript amplified from pale tremor RNA lacks exons 19 to 23 from the 3′ end of the gene (Fig. 1d). Hybridization of a polyA+ northern blot with a complementary DNA probe containing exons 8 to 15 identified a wild-type transcript of 3.3 kilobases (kb) that is not present in messenger RNA from pale tremor mice (Fig. 1e). No abnormal transcripts were detected in the mutant, even when the exposure time was increased from 3 to 63 h (not shown). We were able to amplify exons 19 to 23 from genomic DNA, eliminating the possibility of a genomic deletion (Supplementary Fig. 2a).

To identify the genomic mutation, we examined the structure of intron 18 using PCR. A wild-type product of 0.65 kb and a mutant product of 6.2 kb were amplified from the 3′ end of intron 18 (Fig. 1f). The sequence of the mutant product (GenBank DQ813648) contains a 5,547 base pair (bp) insert with 99% sequence identity to mouse retrotransposon *ETn2β* (GenBank Y17106). The transposon is inserted 384 bp upstream of exon 19, in the same orientation as the gene (Fig. 1h), and is flanked by a duplication of the hexanucleotide CCCCTG, characteristic of *ETn2β* insertions[3]. The mutant allele can be detected by PCR with a primer in the long terminal repeat (Fig. 1g). The background strains do not contain the *ETn2β* element (data not shown), indicating that mutation of the pale tremor gene is a result of transposon insertion. The data are consistent with abnormal splicing from exon 18 of *Fig4* to one of the cryptic splice acceptor sites in the *ETn2β* element[3], generating a transcript of very low abundance that is detected by RT–PCR but is below the sensitivity of the polyA+ northern blot.

RT–PCR of tissues from wild-type mice demonstrated widespread expression of *Fig4* (Supplementary Fig. 2b), consistent with public expressed-sequence-tag and microarray databases. *In situ* hybridization data demonstrate distribution of the transcript throughout the brain (http://www.brainatlas.org/aba/ and unpublished observations, C.Y.C. and M.H.M.). The human orthologue, KIAA0274, is located on human chromosome 6q21. The mutated protein is most closely related to the yeast *Saccharomyces cerevisiae* SAC-domain phosphatase Fig4, with overall amino acid sequence identity of 35% and similarity of 66%. The SAC domain with the active site motif CX$_5$R(S/T) is characteristic of phosphatases with specificity for phosphoprotein or phospholipid substrates[4], and exhibits 44% sequence identity (191/435 amino acids; Supplementary Fig. 1b). The

[1]Department of Human Genetics, [2]Life Sciences Institute, [3]Department of Cellular and Developmental Biology, and [4]Department of Neurology, University of Michigan, Ann Arbor, Michigan 48109, USA. [5]Departments of Molecular and Human Genetics, [6]Pediatrics, and [7]Neurology, Baylor College of Medicine, [8]Texas Children's Hospital, Houston, Texas 77030, USA. [9]Department of Neurology, Wayne State University School of Medicine, Detroit, Michigan 48201, USA. [10]John D. Dingle VA Medical Center, Detroit, Michigan 48201, USA.

four other mammalian genes with SAC domains (*SYNJ1*, *SYNJ2*, *INPP5F* and *SAC1*) differ from Fig4 at other domains[4], indicating that mouse *A530089I17Rik* and human KIAA0274 are homologues of yeast Fig4.

Yeast Fig4 is localized to the vacuolar membrane and is required for both generation and turnover of PtdIns(3,5)P$_2$ (ref. 5). Yeast Fig4

exhibits lipid phosphatase activity towards the 5-phosphate residue of PtdIns(3,5)P$_2$ (ref. 6), and also appears to activate the Fab1 kinase that synthesizes PtdIns(3,5)P$_2$ from phosphatidylinositol-3-phosphate, possibly by dephosphorylating the kinase or one of its regulators[5]. As a result, deletion of yeast Fig4 reduces rather than increases the intracellular concentration of PtdIns(3,5)P$_2$ (ref. 7),

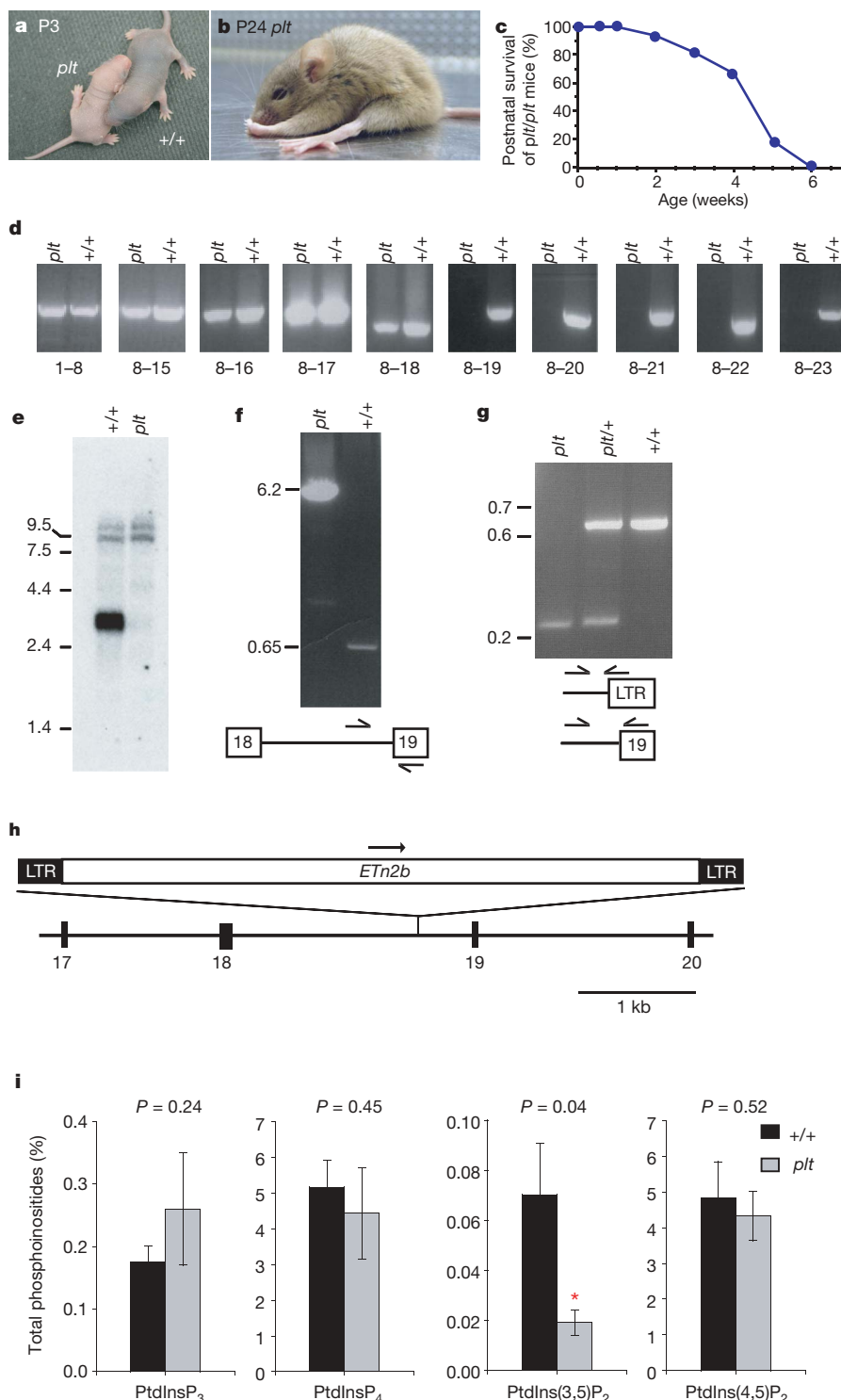**Figure 1 | Phenotypes of homozygous pale tremor mice. a,** Diluted pigmentation of the pale tremor mouse (*plt*) compared with the wild-type (+/+) mouse. **b,** Abnormal limb postures. **c,** Juvenile lethality of F$_2$ mice from the CAST/Ei cross (*n* = 50). **d,** RT–PCR of *Fig4* transcript from brain with primers in indicated exons. **e,** Northern blot containing 3 μg of brain polyA$^+$ RNA, isolated at P7 before extensive neurodegeneration, using a 1 kb *Fig4* cDNA probe (exons 8–15). RNA integrity and equal loading of samples is indicated by the intensity of minor bands (9–10 kb). **f,** Long-range PCR of genomic DNA with primers in intron 18 and exon 19. **g,** Three-primer genotyping assay for Fig4 produces 646 bp wild-type and 245 bp *Fig4*^{paletremor} products. LTR, long terminal repeat. Molecular weight markers in **e–g** are given in kb. **h,** *ETn2β* retrotransposon in intron 18 of *Fig4*. **i,** Altered abundance of PtdIns(3,5)P$_2$ in cultured fibroblasts. *Significant difference at *P* < 0.05. Data are expressed as mean ± s.d. (*n* = 6).

leading to defects in vacuole fission, formation of enlarged vacuoles and impaired retrograde traffic to the late endosome[2,8,9]. Knockdown of mammalian Fab1/PIKfyve kinase causes a similar defect in retrograde endosome traffic and enlarged vacuoles[10].

Analysis of phosphoinositides from cultured fibroblasts of pale tremor mice demonstrated a three-fold reduction in PtdIns(3,5)P$_2$ ($P = 0.04$), with no change in three other phosphoinositides (Fig. 1i). Enlarged cytoplasmic vacuoles accumulate in 40% of cultured fibroblasts from pale tremor mice (174/435) compared with 5% of wild-type cells (22/403) (Supplementary Fig. 3a–d). These vacuoles stain positively for LAMP-2 (Supplementary Fig. 3e–g), indicating that they represent late-stage endosomes. The altered levels of PtdIns(3,5)P$_2$ demonstrate conserved enzymatic function of FIG4 from yeast to mouse, whereas the enlarged vacuoles demonstrate a conserved cellular role in regulation of the size of late endosomes.

In vivo loss of FIG4 results in a striking pattern of selective neurodegeneration. Extensive loss of neurons from sensory and autonomic ganglia is evident during the neonatal period (Fig. 2a–d and Supplementary Fig. 4a). The presence of neurons with enlarged cytoplasmic vacuoles suggests that vacuole accumulation precedes cell loss (Fig. 2, insets). Spinal motor neurons exhibit normal morphology at three weeks (Fig. 2e–f) but contain vacuoles at six weeks of age (Supplementary Fig. 5).

Peripheral nerves are also affected. Cross-sections of the sciatic nerve reveal substantial reduction in the number of large-diameter myelinated axons in the mutant (Fig. 3a, b). Nerve conduction velocity and amplitude of compound muscle action potentials are reduced (Fig. 3c, d), consistent with the axonal loss visible in the sections of the sciatic nerve and the pathological abnormalities in the motor neurons. There was no response when recording from tail sensory fibres, consistent with the severe loss of sensory neurons from the dorsal root ganglia shown in Fig. 2a.

In the brain, neuronal loss in the thalamus, pons, medulla and deep cerebellar nuclei is visible at one week of age (Supplementary Fig. 6). By three weeks of age there is additional loss of neurons from layers four and five of the cortex, the deep layers of the superior and inferior colliculus and the olfactory bulb (Fig. 2g). Some neuronal cell bodies contain enlarged vacuoles (Fig. 2h); these vacuoles are not stained by Oil Red O (for lipid) or periodic acid Schiff (for carbohydrate) (not shown). Hippocampus, cerebellar cortex and cerebral cortex layers one, two, three and six remain relatively unaffected throughout the course of the disease, although under culture conditions hippocampal neurons become vacuolated (Supplementary Fig. 4b, c). The abnormal gait and motor coordination in the pale tremor mutant may be accounted for by abnormal proprioception owing to degeneration of dorsal root ganglia neurons, in combination with degeneration of neurons from brain regions directly involved in motor control (layer five of the cortex, thalamus and deep cerebellar nuclei).

Abnormalities are visible in skin and spleen. Pigment-containing hair follicles are greatly reduced in number and the few pigmented hairs contain clumped melanosomes (Supplementary Fig. 7), similar to mouse mutants with defects in lysosome–melanosome biogenesis[11]. There is extensive cell loss in the spleen (Supplementary Fig. 8). White blood cell profiles are normal, and liver, kidney and testis appear normal when observed by light microscopy.

The clinical and pathological features of peripheral neuropathy in pale tremor mice resemble some types of Charcot–Marie–Tooth (CMT) disorder[12,13]. We tested *FIG4* as a candidate gene by screening 95 individuals diagnosed with CMT disorder but lacking mutations in known genes[13]. The 23 coding exons of *FIG4* were amplified from genomic DNA, screened by heteroduplex analysis and sequenced. Patient BAB1079 has a severe, early onset disorder. We identified the protein truncation mutation F98fsX102 in exon 4 and the missense mutation I41T in exon 2 of *FIG4* (Fig. 4a). F98fsX102 truncates
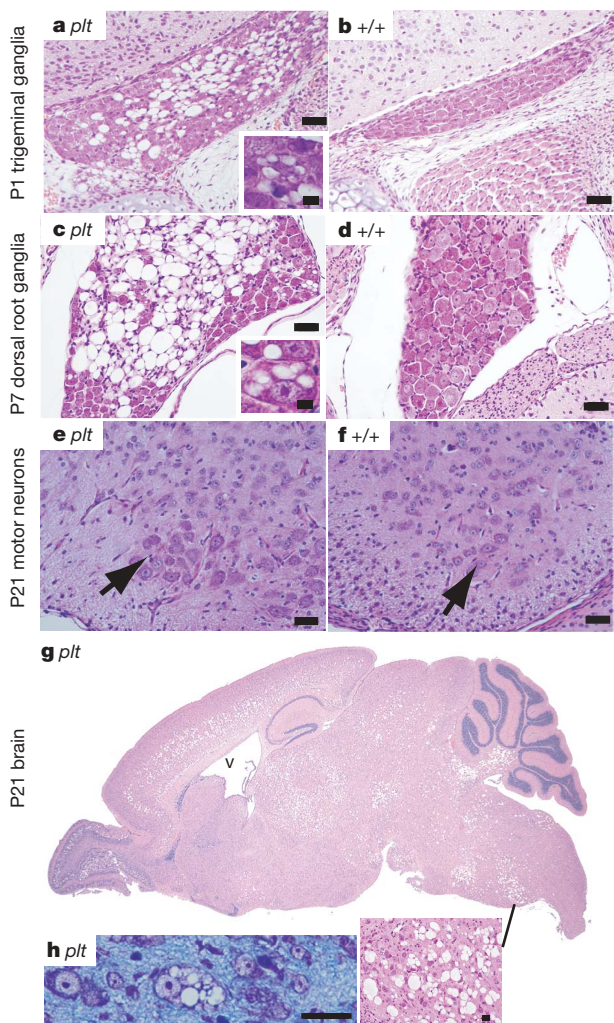

Figure 2 | Neuropathology in pale tremor mice. a, b, Trigeminal ganglia; **c, d,** dorsal root ganglia from lumbar region. Insets reveal cytoplasmic vacuoles. (Superior cervical ganglia have a similar appearance; see Supplementary Fig. 4a.) **e, f,** Spinal cord ventral horn. Arrow, motor neuron cell body. **g,** Sagittal section of the brain of a pale tremor mouse (for wild-type control lacking degeneration see Supplementary Fig. 6). V, enlarged ventricle. **h,** Neuronal cell bodies from regions of degeneration in P7 brain. Scale bars: **a–f,** 25 μm; insets in **a** and **c,** 12.5 μm; panel **h,** 25 μm.
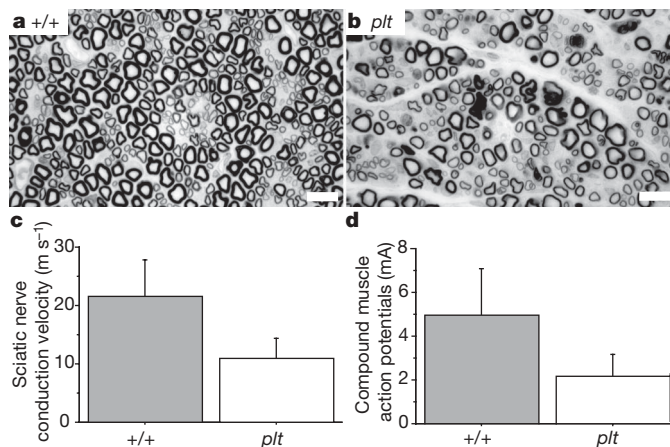

Figure 3 | Pathological abnormalities in peripheral nerves. a, b, Semi-thin sections of sciatic nerve prepared as described[27], demonstrating reduced density of large-diameter myelinated axons in the pale tremor mutant. Scale bars, 10 μm. **c, d,** Reduced sciatic nerve conduction velocity (mutant, 11.0 ± 3.4 m s^{-1}; wild type, 21.5 ± 6.3 m s^{-1}) and reduced amplitude of compound muscle action potentials (mutant, 2.2 ± 1.0 mA; wild type, 5.0 ± 2.1 mA) (mean ± s.d., $n = 6$).

the protein within the SAC domain and is likely to be a loss-of-function allele. Pedigree analysis demonstrated autosomal recessive inheritance. Each mutation was inherited from a heterozygous parent (Fig. 4b). Two heterozygous carriers of F98fsX102 are unaffected, indicating that *FIG4* is not haploinsufficient. Heterozygous pale tremor mice at 18 months of age are also unaffected.

Another patient, BAB1372, was found to be a compound heterozygote with a nonsense mutation R183X in exon 6 together with I41T (Fig. 4a). R183X was inherited from the patient's father (Fig. 4b). The mother is an obligate carrier of I41T, and the patient has an affected sibling (BAB1373) that also inherited both mutations (Fig. 4b). Both siblings have severe disease: BAB1372 is functionally quadriplegic and BAB1373 is wheelchair-bound although retaining normal use of his arms. Both have slow nerve conduction velocities. A sural nerve biopsy for BAB1373 demonstrated profound axonal loss, thinly myelinated nerve fibres and evidence of de- and re-myelination.

Two additional patients, BAB1161 and BAB1369, carry unique truncation mutations together with I41T (Fig. 4a). Both patients developed disease by the age of five and demonstrate reduced nerve conduction velocity (2–7 m s^{-1}, compared with normal values of 40–50 m s^{-1}). One patient had motor developmental delay consistent with Dejerine–Sottas neuropathy.

It is remarkable that four unrelated Caucasian patients carry the same missense mutation. No additional coding or splice site variants were detected in these patients when all 23 exons of *FIG4* were sequenced. Isoleucine 41 is located upstream of the SAC phosphatase domain and is evolutionarily invariant in *FIG4* from yeast, invertebrates and vertebrates (Fig. 4c, d). We sequenced exon 2 from 295 neurologically normal Caucasian controls but did not identify carriers of this variant. The observed allele frequencies were 0/590 in controls and 4/190 in CMT patients ($P = 0.003$). The four patients carry I41T on the same 15 kb haplotype, defined by single nucleotide polymorphisms rs3799845 (G), rs2025249 (C) and rs7764711 (G) (haplotype frequency 0.29; disequilibrium coefficient, $D' = 1$), consistent with inheritance of a common ancestral mutant allele (see data in Supplementary Fig. 9). The evidence suggests that I41T is a rare allele causing partial loss of function that is pathogenic in combination with a null allele of *FIG4*.

To assess the functional effect of the I41T allele, we tested the corresponding mutation (I59T) in yeast. Wild-type and mutant cDNAs were transformed into the yeast strain fig4Δ lacking functional Fig4 (ref. 7). Transformation with the empty vector did not correct the enlarged vacuoles in fig4Δ, which reflects the slightly reduced levels of PtdIns(3,5)P$_2$ that result from the absence of Fig4 (ref. 7) (Fig. 5a). The vacuolar enlargement was corrected to a comparable extent by wild-type and mutant Fig4, indicating that, under basal conditions, cells expressing Fig4^{I59T} produce normal levels of PtdIns(3,5)P$_2$ (Fig. 5a). The ability of the mutant to activate Fab1/PIKfyve kinase was tested by treatment with hyperosmotic shock as previously described[5,7]. In cells expressing wild-type Fig4, hyperosmotic shock produces a transient tenfold increase in intracellular PtdIns(3,5)P$_2$ concentration owing to activation of Fab1/PIKfyve kinase (Fig. 5b). In cells expressing the mutant, a partial fourfold increase was observed, demonstrating impaired activation of Fab1/PIKfyve kinase. It is not clear whether the phosphatase activity of the mutant is also impaired, owing to the low levels of PtdIns(3,5)P$_2$.

The data presented here demonstrate that mutation of *FIG4* is responsible for peripheral neuronopathy in human patients. We propose the designation CMT4J for this disorder, based on the recessive inheritance pattern. Phosphoinositide signalling has previously been implicated in CMT disorder types 4B1, 4B2 and 4H (ref. 14–19) (see Supplementary Discussion). Other genes that function in vesicle trafficking, such as *RAB7A* and *DNM2* in human[20,21] and *Vps54* in mouse[22], are associated with inherited neuropathies.

The pale tremor mutant provides the first evidence regarding the functional role of mammalian *Fig4*. The results demonstrate a conserved biochemical function in metabolism of PtdIns(3,5)P$_2$, a conserved cellular role in regulation of endosomal vesicles, and an unexpected role in neuronal survival. The molecular basis for the differential sensitivity of neuronal subtypes to loss of this widely expressed gene is unclear. Neuronal dependence on FIG4 may be related to the role of endosomal vesicles in delivering membrane components to dendritic spines during long-term potentiation[23]. The pale tremor mouse will generate insights into PtdIns(3,5)P$_2$ signalling in
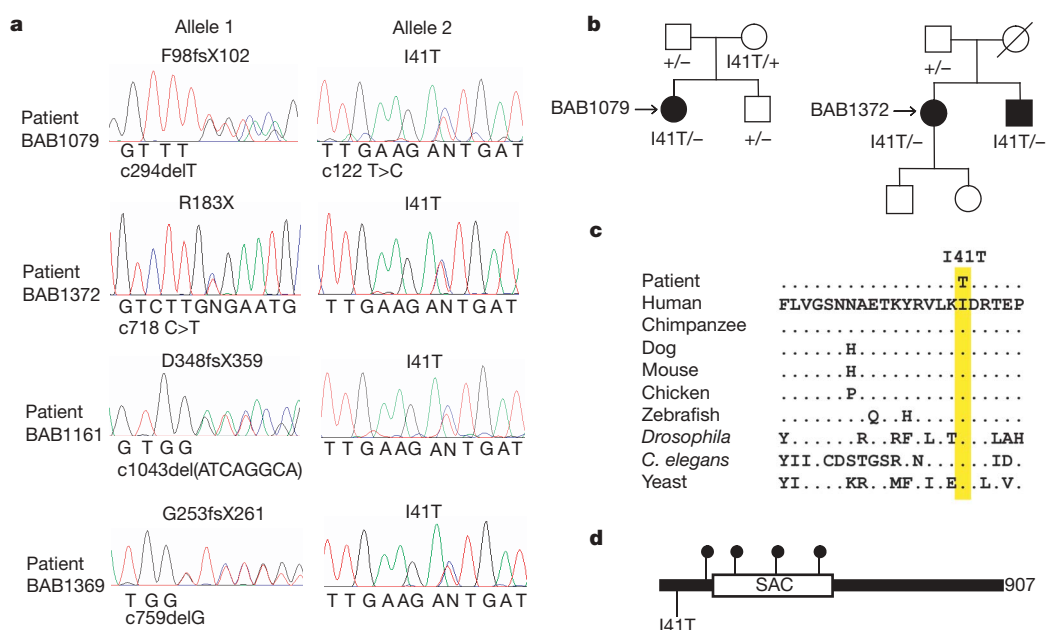

Figure 4 | Mutations of *FIG4* in patients with CMT disorder. a, Sequencing chromatographs for four unrelated patients with CMT disorder. Nucleotide mutations are indicated on the *x*-axis of each chromatograph numbered from +1 for the first codon. **b**, Inheritance of mutant alleles in two pedigrees. Circle, female; square, male; open symbol, unaffected; filled symbol, affected. Patient BAB1079 is a compound heterozygote for protein truncation mutation F98fsX102 and missense mutation I41T. Patients BAB1372 and BAB1373 are compound heterozygotes for nonsense mutation R183X and missense mutation I41T. **c**, Residue isoleucine 41 is evolutionarily invariant in FIG4 from vertebrates, invertebrates and yeast. **d**, Location of CMT mutations in the FIG4 protein. Solid circles, protein truncation mutations. 907 amino acids are present per protein.
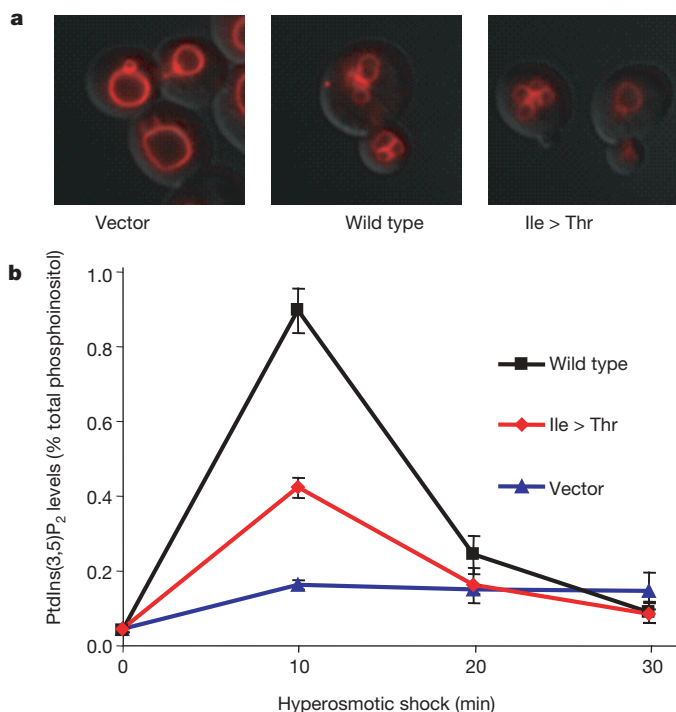
a

Vector — Wild type — Ile > Thr

b

Figure 5 | Yeast Fig4$^{Ile > Thr}$ is defective in activation of kinase Fab1/PIKfyve. The fig4Δ yeast strain lacking endogenous Fig4 (refs 5, 7) was transformed with empty vector (vector), wild-type myc-Fig4 or myc-Fig4 containing I59T corresponding to human I41T (mutant). **a,** Yeast were labelled with FM4-64 to assess vacuole volume, an indicator of basal levels of PtdIns(3,5)P_2. **b,** Time course of PtdIns(3,5)P_2 levels after hyperosmotic shock (measured in minutes after change to hyperosmotic medium) reflects activation of Fab1 kinase. The mutant protein exhibits impaired activation at 10 min compared with the wild type ($P = 0.004$; mean ± s.d., $n = 3$). A western blot demonstrating comparable expression of wild-type and mutant constructs is presented in Supplementary Fig. 10.

neurons and provide an animal model for CMT4J and related human neuropathies.

METHODS SUMMARY

Mutation of the pale tremor gene arose on a mixed background derived from inbred mouse strains 129/Ola, C57BL/6J, C3H and SJL (ref. 24). The mutant allele is designated $Fig4^{paletremor}$. Total RNA was isolated from brain of P7 mice before extensive neurodegeneration. Primary fibroblasts were cultured from mouse tail biopsies treated with collagenase. Phosphoinositides were labelled for 24 h with myo-[2-^{3}H] inositol, extracted and quantified by high-performance liquid chromatography (HPLC, see Methods). The cohort of unrelated patients with CMT disorder was previously described[13]. The 23 coding exons of *FIG4* were screened by heteroduplex analysis[25] and products with abnormal mobility were sequenced. For the four individuals with identified mutations, all of the other exons were sequenced. Genomic DNA from neurological normal control individuals was obtained from the Coriell Institute (NDPT006 and NDPT009, 96 samples each) and from a collection of 111 subjects older than 60 without personal or family history of neurological disease[26].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Di Paolo, G. & De Camilli, P. Phosphoinositides in cell regulation and membrane dynamics. *Nature* **443**, 651–657 (2006).
2. Michell, R. H., Heath, V. L., Lemmon, M. A. & Dove, S. K. Phosphatidylinositol 3,5-bisphosphate: metabolism and cellular functions. *Trends Biochem. Sci.* **31**, 52–63 (2006).
3. Maksakova, I. A. *et al.* Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.* **2**, e2 (2006).
4. Hughes, W. E., Cooke, F. T. & Parker, P. J. Sac phosphatase domain proteins. *Biochem. J.* **350**, 337–352 (2000).
5. Duex, J. E., Tang, F. & Weisman, L. S. The Vac14p–Fig4p complex acts independently of Vac7p and couples PI3,5P_2 synthesis and turnover. *J. Cell Biol.* **172**, 693–704 (2006).
6. Rudge, S. A., Anderson, D. M. & Emr, S. D. Vacuole size control: regulation of PtdIns(3,5)P_2 levels by the vacuole-associated Vac14–Fig4 complex, a PtdIns(3,5)P_2-specific phosphatase. *Mol. Biol. Cell* **15**, 24–36 (2004).
7. Duex, J. E., Nau, J. J., Kauffman, E. J. & Weisman, L. S. Phosphoinositide 5-phosphatase Fig 4p is required for both acute rise and subsequent fall in stress-induced phosphatidylinositol 3,5-bisphosphate levels. *Eukaryot. Cell* **5**, 723–731 (2006).
8. Bonangelino, C. J. *et al.* Osmotic stress-induced increase of phosphatidylinositol 3,5-bisphosphate requires Vac14p, an activator of the lipid kinase Fab1p. *J. Cell Biol.* **156**, 1015–1028 (2002).
9. Gary, J. D. *et al.* Regulation of Fab1 phosphatidylinositol 3-phosphate 5-kinase pathway by Vac7 protein and Fig4, a polyphosphoinositide phosphatase family member. *Mol. Biol. Cell* **13**, 1238–1251 (2002).
10. Rutherford, A. C. *et al.* The mammalian phosphatidylinositol 3-phosphate 5-kinase (PIKfyve) regulates endosome-to-TGN retrograde transport. *J. Cell Sci.* **119**, 3944–3957 (2006).
11. Marks, M. S. & Seabra, M. C. The melanosome: membrane dynamics in black and white. *Nature Rev. Mol. Cell Biol.* **2**, 738–748 (2001).
12. Schroder, J. M. Neuropathology of Charcot–Marie–Tooth and related disorders. *Neuromolecular Med.* **8**, 23–42 (2006).
13. Szigeti, K., Garcia, C. A. & Lupski, J. R. Charcot–Marie–Tooth disease and related hereditary polyneuropathies: molecular diagnostics determine aspects of medical management. *Genet. Med.* **8**, 86–92 (2006).
14. Begley, M. J. *et al.* Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase. *Proc. Natl Acad. Sci. USA* **103**, 927–932 (2006).
15. Bolino, A. *et al.* Disruption of *Mtmr2* produces CMT4B1-like neuropathy with myelin outfolding and impaired spermatogenesis. *J. Cell Biol.* **167**, 711–721 (2004).
16. Bolino, A. *et al.* Charcot–Marie–Tooth type 4B is caused by mutations in the gene encoding myotubularin-related protein-2. *Nature Genet.* **25**, 17–19 (2000).
17. Bonneick, S. *et al.* An animal model for Charcot–Marie–Tooth disease type 4B1. *Hum. Mol. Genet.* **14**, 3685–3695 (2005).
18. Senderek, J. *et al.* Mutation of the *SBF2* gene, encoding a novel member of the myotubularin family, in Charcot–Marie–Tooth neuropathy type 4B2/11p15. *Hum. Mol. Genet.* **12**, 349–356 (2003).
19. Stendel, C. *et al.* Peripheral nerve demyelination caused by a mutant Rho GTPase guanine nucleotide exchange factor, frabin/FGD4. *Am. J. Hum. Genet.* (in the press); preprint at http://www.journals.uchicago.edu/AJHG/journal/preprints/AJHG44688.preprint.pdf (2007).
20. Verhoeven, K. *et al.* Mutations in the small GTP-ase late endosomal protein RAB7 cause Charcot–Marie–Tooth type 2B neuropathy. *Am. J. Hum. Genet.* **72**, 722–727 (2003).
21. Zuchner, S. *et al.* Mutations in the pleckstrin homology domain of dynamin 2 cause dominant intermediate Charcot–Marie–Tooth disease. *Nature Genet.* **37**, 289–294 (2005).
22. Schmidt-John, T. *et al.* Mutation of *Vps54* causes motor neuron disease and defective spermiogenesis in the wobbler mouse. *Nature Genet.* **37**, 1213–1215 (2005).
23. Park, M. *et al.* Plasticity-induced growth of dendritic spines by exocytic trafficking from recycling endosomes. *Neuron* **52**, 817–830 (2006).
24. Adamska, M., Billi, A. C., Cheek, S. & Meisler, M. H. Genetic interaction between *Wnt7a* and *Lrp6* during patterning of dorsal and posterior structures of the mouse limb. *Dev. Dyn.* **233**, 368–372 (2005).
25. Escayg, A. *et al.* Mutations of *SCN1A*, encoding a neuronal sodium channel, in two families with GEFS+2. *Nature Genet.* **24**, 343–345 (2000).
26. Rainier, S., Sher, C., Reish, O., Thomas, D. & Fink, J. K. De novo occurrence of novel *SPG3A*/atlastin mutation presenting as cerebral palsy. *Arch. Neurol.* **63**, 445–447 (2006).
27. Li, J. *et al.* Major myelin protein gene (P0) mutation causes a novel form of axonal degeneration. *J. Comp. Neurol.* **498**, 252–265 (2006).
28. Kohrman, D. C., Harris, J. B. & Meisler, M. H. Mutation detection in the *med* and *med^J* alleles of the sodium channel *Scn8a*. Unusual splicing due to a minor class AT–AC intron. *J. Biol. Chem.* **271**, 17576–17581 (1996).

nature

METHODS

Animals. For genetic mapping, *plt*/+ heterozygotes were crossed with strain CAST/Ei (Jackson Laboratory). Experiments were carried out on F_2 and F_3 mice from the mapping cross. This research was approved by the University of Michigan Committee on Use and Care of Animals. Animals were housed and cared for in accordance with NIH guidelines.

Genotyping and markers. Genotyping was carried out using microsatellite markers from public databases as well as novel microsatellite markers designed from mouse genomic sequence (http://www.ensemble.org). D10Umi13 was amplified with the forward primer 5′-CCACCACATCAACAGGCTCACAGG and the reverse primer 5′-AATGCAACCGTGACACAAGTACAC. PCR was carried out with the PCR core kit (Qiagen). PCR products were separated on 6% acrylamide gels and stained with ethidium bromide. The pale tremor mutation is genotyped by PCR with a forward primer in intron 18 (5′-CGTATGAATT-GAGTAGTTTTGATG) and two reverse primers, one in the proximal LTR of the inserted *ETn2β* element (5′-GCTGGGGGAGGGGAGACTACACAG) and one in exon 19 (5′-ATGGACTTGGATCAATGCCAACAG).

RT–PCR. Total RNA was isolated from brain of P7 mice, before extensive neuro-degeneration. cDNA was synthesized using the First Strand cDNA Synthesis Kit (Invitrogen). RT–PCR was carried out with the PCR Core Kit (Qiagen). Long-range PCR was performed with the Expand Long Template PCR System (Roche). The northern blot with 3 μg of polyA$^+$ RNA was prepared as previously described[28]. The hybridization probe, a 1 kb RT–PCR product containing exons 8 to 15, was labelled with two radiolabelled nucleotide triphosphates.

Histology. Tissues were sectioned and stained at HistoServ. Fast blue/eosin staining was carried out in the Department of Pathology, University of Michigan. Light microscopy was performed on an Olympus BX-51 microscope and DP50 camera. Sciatic and femoral nerves were sectioned and stained with osmium for electron microscopy as previously described[27]. Skin whole mounts were prepared from P10 mice with the guidance of A. Dlugosz. The commercial depilatory Nair was applied to the dorsal surface for five minutes followed by washing with warm water to remove hair. The skin was dissected and superficial fascia removed. Follicles were visualized on a standard dissecting microscope with transmitted light.

Neurophysiology. Nerve conduction velocities were recorded from affected pale tremor mice and littermate controls. Mice were anaesthetized with a ketamine/xylazine solution and placed under a heating lamp to maintain body temperature at 32 °C. Recordings were obtained using a Nicolet VikingQuest portable system and Nicolet disposable EEG needles. Tail sensory responses were obtained by stimulating proximally over a 3 cm region. Sciatic nerve motor velocities were obtained by stimulating distally at the sciatic notch and proximally at the knee.

Cell culture and immunofluorescence. Primary fibroblasts were cultured from mouse tail biopsies treated with collagenase. Cells were plated in DMEM with 10% fetal bovine serum, and maintained at 37 °C with 5% CO_2 for up to three passages. For immunofluorescence, 100,000 cells were seeded on polylysine-coated cover slips in 35 mm dishes. We thank T. August and the Developmental Studies Hybridoma Bank for monoclonal antibody to LAMP-2, developed under the auspices of the NICHD and maintained by the Department of Biological Sciences, University of Iowa. For labelling with antibody, cells were fixed with ice-cold methanol at −20 °C for 5 min and blocked with 2% goat serum. Antibodies were applied for 1 h in PBS with 2% serum at room temperature and detected with Alexa Fluor 488 donkey anti-rabbit (Molecular Probes) or Alexa Fluor 594 goat anti-rat (Molecular Probes). Cells were visualized on a DeltaVision Deconvolution microscope system (Applied Precision).

Hippocampal neurons were cultured with glial conditioned media. Neurons were visualized with a Nikon TE2000 microscope.

Phosphoinositide assays. Fibroblast phosphoinositides were labelled with myo-[2-^{3}H] inositol, extracted and quantified by HPLC. Mouse fibroblasts from the first passage were grown in 100 mm dishes to 60–70% confluency. The culture was rinsed with PBS and starved for 12 h in inositol-free DMEM (Tissue Culture Support Centre, Washington University) supplemented with 5 μg ml^{-1} transferrin, 5 μg ml^{-1} insulin and 10% dialysed fetal bovine serum. The medium was replaced with labelling medium (inositol-free DMEM containing 5 μg ml^{-1} transferrin, 20 mM HEPES and 50 μCi myo-[2-^{3}H] inositol, GE Healthcare). After 24 h, the culture was treated with 0.6 ml of 4.5% (v/v) perchloric acid for 15 min, scraped off the plate, and spun down at 12,000g for 10 min at 4 °C. The pellet was washed with 0.1 M EDTA once and resuspended in 50 μl deionized water. To deacylate the lipids, samples were transferred to a glass vial, mixed with 1 ml methanol/40% methylamine/*n*-butanol (4:4:1, v/v), and incubated at 55 °C for 1 h. The resulting samples were vacuum-dried, resuspended in 0.3 ml water and extracted twice with an equal volume of *n*-butanol/ethyl ether/formate (20:4:1, v/v). The aqueous phase was vacuum dried and resuspended in 20 μl water.

For separation of all isoforms of the glycerophosphoinositides by HPLC, two different elution gradients were used at 1 ml min^{-1} flow rate. (Pump A, H_2O; pump B, 1 M $(NH_4)_2HPO_4$, pH 3.8.) Gradient 1: 0% B for 5 min; 0–2% B over 15 min; 2% B for 80 min; 2–12% B over 20 min; 12% B for 20 min; 12–80% B over 40 min; 80% B for 20 min; 80–0% B over 5 min. To separate glyceroPtdIns(3,4)P_2 from glyceroPtdIns(3,5)P_2, a longer gradient (gradient 2) was used: 0% B for 5 min; 0–2% B over 15 min; 2% B for 80 min; 2–10% B over 20 min; 10% B for 65 min; 10–80% B over 40 min; 80% B for 20 min; 80–0% B over 5 min. The positions of glyceroPtdIns(3)P, glyceroPtdIns(3,5)P_2, glyceroPtdIns(3,4)P_2 and glyceroPtdIns(3,4,5)P_3 were determined by ^{32}P-labelled standards received as gifts from L. Rameh. The positions of glyceroPtdIns(4)P and glyceroPtdIns(4,5)P_2 were confirmed with yeast glycerophosphoinositide extracts.

Human mutation detection. The cohort of unrelated patients with CMT disorder was previously described[13]. The clinical diagnosis was based on clinical examination, electrophysiological studies and, in a few cases, nerve biopsy. All patients received appropriate counselling and gave informed consent approved by the institutional review board. For the initial screen of *FIG4*, each coding exon was amplified and examined by heteroduplex analysis as previously described[25]. The patient mutations were identified by sequencing products exhibiting abnormal mobility. Subsequently, the 23 exons of *FIG4* were completely sequenced from the four individuals carrying variants.

LETTERS

Novel neurotrophic factor CDNF protects and rescues midbrain dopamine neurons *in vivo*

Päivi Lindholm[1], Merja H. Voutilainen[2], Juha Laurén[1]†, Johan Peränen[1], Veli-Matti Leppänen[1], Jaan-Olle Andressoo[1], Maria Lindahl[1], Sanna Janhunen[2]†, Nisse Kalkkinen[1], Tõnis Timmusk[1,3], Raimo K. Tuominen[2] & Mart Saarma[1]

In Parkinson's disease, brain dopamine neurons degenerate most prominently in the substantia nigra[1]. Neurotrophic factors promote survival, differentiation and maintenance of neurons in developing and adult vertebrate nervous system[2,3]. The most potent neurotrophic factor for dopamine neurons described so far is the glial-cell-line-derived neurotrophic factor (GDNF)[4]. Here we have identified a conserved dopamine neurotrophic factor (CDNF) as a trophic factor for dopamine neurons. CDNF, together with its previously described vertebrate and invertebrate homologue the mesencephalic-astrocyte-derived neurotrophic factor[5], is a secreted protein with eight conserved cysteine residues, predicting a unique protein fold and defining a new, evolutionarily conserved protein family. CDNF (Armetl1) is expressed in several tissues of mouse and human, including the mouse embryonic and postnatal brain. *In vivo,* CDNF prevented the 6-hydroxydopamine (6-OHDA)-induced degeneration of dopaminergic neurons in a rat experimental model of Parkinson's disease. A single injection of CDNF before 6-OHDA delivery into the striatum significantly reduced amphetamine-induced ipsilateral turning behaviour and almost completely rescued dopaminergic tyrosine-hydroxylase-positive cells in the substantia nigra. When administered four weeks after 6-OHDA, intrastriatal injection of CDNF was able to restore the dopaminergic function and prevent the degeneration of dopaminergic neurons in substantia nigra. Thus, CDNF was at least as efficient as GDNF in both experimental settings. Our results suggest that CDNF might be beneficial for the treatment of Parkinson's disease.

GDNF delivery into the putamen of parkinsonian patients has resulted in significant clinical improvement[6,7]. However, recent data showed low clinical benefit of GDNF and raised serious safety issues[8]. Therefore, it is important to search for new neurotrophic factors that could slow down or reverse the progression of neuronal degeneration and that are well tolerated.

Here we describe CDNF, a new neurotrophic factor for dopaminergic neurons (Fig. 1a). CDNF is a vertebrate specific paralogue of the recently identified human mesencephalic-astrocyte-derived neurotrophic factor (MANF)[5]. Using bioinformatics, we identified several human and mouse expressed sequence tag (EST) complementary DNA sequences homologous to *MANF*. We cloned full-length *CDNF* cDNAs from human and mouse brain RNA by using PCR with reverse transcription (RT–PCR). In public databases *MANF* and *CDNF* genes are named as *ARMET* (arginine-rich, mutated in early stage tumors) and *ARMETL1* (ARMET-like 1), respectively. However, the name '*ARMET*' seems misleading because only human

MANF, but not genes homologous to *MANF* or *CDNF* in other species, including mouse, contains an amino-terminal arginine-rich sequence. In human *MANF* the arginine-rich region is considered as a non-translated sequence[5].

CDNF and MANF proteins form a novel MANF family of conserved secreted factors with eight cysteine residues of similar spacing (Fig. 1a, b), suggesting a unique protein fold. Human and mouse *CDNF* messenger RNAs encode 187 amino acid proteins with predicted signal peptides and apparently no pro-sequence. Predicted secondary structure of human and mouse CDNFs, like that of MANF, is dominated by α-helices (Fig. 1a). Analyses of genome and EST sequences from various organisms suggest that vertebrates, including the frog *Xenopus tropicalis* and zebrafish *Danio rerio*, have orthologous genes for both *CDNF* and *MANF*, whereas the invertebrates, including the nematode *Caenorhabditis elegans* and fruitfly *Drosophila melanogaster* have a single ancestral *MANF/CDNF* gene orthologue (Supplementary Fig. 1). Human CDNF shows 59% amino acid identity with human MANF, 49% identity with *D. melanogaster* and 46% identity with *C. elegans* MANF proteins (Supplementary Table 1).

To study the secretion and post-translational modification of CDNF protein, we cloned human CDNF full-length cDNA into a mammalian expression vector. CDNF was efficiently secreted from transiently transfected HEK 293T cells (Fig. 1c). The signal-sequence cleavage site of human CDNF secreted from COS-7 African green monkey fibroblast cells was between residues 26 (glycine) and 27 (glutamine) (Fig. 1a). Electrospray ionization–mass spectrometry (ESI–MS) analysis revealed that the eight cysteine residues formed three or four intramolecular disulphide bonds in the tertiary structure of the mature CDNF protein. To obtain large quantities of CDNF protein, which is needed to assess its function *in vivo*, we produced recombinant human CDNF using a baculovirus expression system in Sf9 cells. CDNF was secreted into the culture medium, from which it was purified to at least 90% purity (Fig. 1d). Purified CDNF was analysed by amino-terminal sequencing and ESI–MS, which revealed that CDNF produced in Sf9 cells was not glycosylated, and contained four intramolecular disulphide bridges.

Cdnf mRNA expression was characterized in developing mouse brain and in various adult mouse tissues by RT–PCR. In the brain, *Cdnf* transcripts were detected in embryonic and postnatal stages (Fig. 2a), and in adult brain regions, including the striatum (Fig. 2b). In the midbrain, *Cdnf* mRNA was expressed from embryonic stages to the adult (Fig. 2c, b). High levels of *Cdnf* mRNA were observed in the adult mouse heart, skeletal muscle and testis

[1]Institute of Biotechnology, P.O. Box 56, Viikki Biocenter, University of Helsinki, FIN-00014, Helsinki, Finland. [2]Division of Pharmacology and Toxicology, Faculty of Pharmacy, P.O. Box 56, Viikki Biocenter, University of Helsinki, FIN-00014, Helsinki, Finland. [3]Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 19086 Tallinn, Estonia. †Present addresses: Department of Neurology, Yale University School of Medicine, P.O. Box 208018, New Haven, Connecticut 06520, USA (J.L.); Department of Pharmacology, CARIM, University of Maastricht, P.O. Box 616, Maastricht, 6200 MD, The Netherlands (S.J.).

(Fig. 2d). In agreement with the results from mouse, a wide expression of *CDNF* transcripts was detected in the human brain and non-neuronal tissues (Supplementary Fig. 2a, b). Interestingly, the levels of *CDNF* transcripts were relatively high in the corpus callosum and optic nerve, which are devoid of neuronal cell bodies and contain mostly axonal projections and oligodendrocytes.

In situ hybridization analysis was used to study the cellular localization of *Cdnf* mRNA in mouse brain. At postnatal stage P10, *Cdnf* labelling was detected in several brain regions (Supplementary Fig. 3a–l) and was found to be most intense in the hippocampus and thalamus (Supplementary Fig. 3e, m, n). *Cdnf* signal was also detected in the striatum and substantia nigra (Supplementary Fig. 3c, i, o, p). In P1 mouse brain, *Cdnf* labelling was weak and widespread (Supplementary Fig. 3q–x) and in accordance with results from P10 brain. In embryonic sections the level of *Cdnf* mRNA was below the detection limit.

To study CDNF protein expression in mammalian tissues, we raised polyclonal antibodies specific for mouse CDNF (Supplementary Fig.

Figure 1 | CDNF is an evolutionarily conserved secreted protein.
a, Alignment of CDNF and MANF amino acid sequences from human (*H.s.*), mouse (*M.m.*), *D. melanogaster* (*D.m.*) and *C. elegans* (*C.e.*). Identical amino acids are marked with an asterisk, physicochemically highly similar with a colon, and similar with a dot. Predicted signal sequences are underlined, and the determined cleavage site in human CDNF is shown by the arrow. The eight conserved cysteines are boxed, and predicted α-helical regions are indicated with a grey background. **b**, Schematic illustration of CDNF and MANF molecules. The cysteines are indicated by bars. Numbers of amino acids are shown. Pre indicates signal sequence that is cleaved off during secretion. **c**, Secretion of native human CDNF (calculated molecular mass, 18 kDa) from transiently transfected HEK 293T cells. Cell lysate and culture medium proteins were analysed by western blotting. The upper band in the cell lysate is most probably pre-CDNF. **d**, Purification steps of recombinant CDNF. Shown is an SDS–PAGE gel stained with Coomassie brilliant blue. Lane 1, concentrated elute from Ni-matrix; lane 2, purified CDNF protein (5 μg).

Figure 2 | Mouse CDNF mRNA and protein are expressed in developing and adult brain and in non-neuronal tissues. RT–PCR analysis of *Cdnf* mRNA in embryonal and postnatal brain (**a**), adult brain regions (**b**), embryonic and postnatal midbrain (**c**) and in adult non-neuronal tissues (**d**). E, embryonic day; P, postnatal day. **e**, A western blot showing CDNF in adult mouse tissues. CDNF appears as a double band in some samples and the upper band probably represents pre-CDNF. **f–n**, Immunohistochemical analysis of CDNF in the adult mouse brain. **f**, CDNF-positive cells in the cerebral cortex (CTX; layer III). **g**, **h**, CDNF antibodies are specific for CDNF and do not cross-react with MANF. **g**, CDNF antibodies pre-incubated with MANF protein (4 μg ml^{-1}) detect CDNF. **h**, No CDNF signal was detected with antibodies pre-incubated with CDNF (4 μg ml^{-1}). **i**, CDNF (green) co-localized with NeuN (red). **j–n**, Cellular localization of CDNF in the CA3 region of hippocampus, substantia nigra (SN), Purkinje cells (Pkj) and locus coeruleus (LC). **l**, CDNF (green) does not co-localize with tyrosine hydroxylase (red) in the substantia nigra. Scale bar in **f** represents 50 μm in **f–h** and **j–n**, and 12 μm in **i**.

4). In agreement with the results of *Cdnf* mRNA expression, relatively high amounts of CDNF protein were detected in the adult mouse heart, skeletal muscle and testis by western blotting (Fig. 2e). Low levels of CDNF were detected in the extracts of whole brains (Fig. 2e).

By immunohistochemical analysis, CDNF-specific signal was detected in the cerebral cortex of adult mouse brain through layers II to VI (Fig. 2f–h). The signal was co-localized with NeuN (Neuna60), a neuronal marker, indicating that CDNF is expressed in cortical neurons (Fig. 2i). In the hippocampus, CDNF was detected in the CA1 to CA3 pyramidal regions (Fig. 2j), and in the granule and polymorph layers of dentate gyrus. In the striatum, weak CDNF staining was detected (not shown). In the substantia nigra, CDNF labelling was detected in solitary cells (Fig. 2k) that did not express tyrosine hydroxylase (TH), a marker for dopaminergic neurons (Fig. 2l). Relatively high levels of CDNF protein were detected in the Purkinje cells of the cerebellum (Fig. 2m) and in regions of the brain stem, including the locus coeruleus (Fig. 2n). In general, CDNF protein was localized in cell somas, which is in agreement with its function as a secretory protein. Immunohistochemical analysis of P1 and P10 brain revealed CDNF signal in several areas, including the hippocampus, thalamus, striatum and substantia nigra (Supplementary Fig. 5). This was in agreement with the results obtained by *in situ* hybridization analysis of *Cdnf* mRNA (Supplementary Fig. 3).

On the basis of the growth factor-like characteristics of CDNF and the data showing that MANF promotes the survival of dopaminergic neurons *in vitro*[5], we decided to test the effect of CDNF on the survival of dopaminergic neurons using a rat unilateral 6-OHDA lesion model of Parkinson's disease[9]. We compared the effect of CDNF with that of GDNF, a neurotrophic factor known to protect dopaminergic neurons[10–13]. The rats were given unilaterally into the striatum either vehicle, CDNF (10 μg) or GDNF (10 μg), 6 h before 6-OHDA (8 μg). In the vehicle-pre-treated 6-OHDA-lesioned rats, amphetamine (2.5 mg kg^{-1}, intraperitoneally (i.p.)) induced a strong ipsilateral turning behaviour both at 2 and 4 weeks post lesion (Fig. 3a, Supplementary Data). In the CDNF-pre-treated rats, the effect of amphetamine was significantly reduced when measured at 2 ($P < 0.01$) but also at 4 weeks ($P < 0.05$) post lesion (Fig. 3a). In the GDNF-pre-treated rats, the rotational effect of amphetamine was significantly reduced at 2 weeks ($P < 0.01$) but not at 4 weeks post lesion. The number of TH-positive cell bodies in the substantia nigra pars compacta (SNpc) and the density of TH-positive fibres in the striatum were significantly higher 4 weeks after the treatments in the 6-OHDA-lesioned rats pre-treated with CDNF (96%, $P < 0.001$, for TH-positive cells; and 75%, $P < 0.001$, for TH-positive fibres) or GDNF (93%, $P < 0.001$, for TH-positive cells; and 69%, $P < 0.01$, for TH-positive fibres), compared with the vehicle-pre-treated rats (65% for TH-positive cells, and 55% for TH-positive fibres; Fig. 3b–j). The protective effect of CDNF was dose-dependent (Fig. 3k, l), and a clear neuroprotection was observed already with a single injection of 3 μg of CDNF. Thus, CDNF prevents the 6-OHDA-induced degeneration of dopaminergic neurons at least as efficiently as GDNF.

Next, we studied the possible neurorestorative activity of CDNF by injecting the animals with 6-OHDA (20 μg per 4 μl) and 4 weeks thereafter with CDNF (10 μg), GDNF (10 μg) or vehicle (controls) into the striatum (Fig. 4a). CDNF and GDNF significantly reduced the amphetamine-induced rotational behaviour as compared with the control group (Fig. 4b). In comparison with the number of TH-positive cells in the intact side of the brain (that is, 100%), the number of TH-positive cells in the lesion side of the SNpc was 26% in the rats treated with vehicle (Fig. 4c). A partial recovery of TH-positive cells was observed in rats treated with CDNF (58%, $P = 0.0629$) and GDNF (57%, $P < 0.05$). These findings suggest that CDNF, similarly to GDNF, restores the dopaminergic neurons after 6-OHDA-lesioning.

In the present work we have described a novel dopaminergic neurotrophic factor CDNF, a member of the highly conserved

MANF protein family. One MANF/CDNF orthologue is found in each of *C. elegans* and *D. melanogaster*[5], organisms in which no neurotrophic factors have been so far described[14]—although the existence of neurotrophic principles and interactions in the *Drosophila* nervous system has become evident[15]. The high homology between the invertebrate MANF and vertebrate CDNF and MANF proteins suggests that the invertebrate MANF functions as a neurotrophic factor.

The expression of *Cdnf* mRNA and protein was detected in postnatal and adult mouse striatum and substantia nigra; that is, in areas where axon terminals and cell somas of the midbrain dopaminergic neurons are located, respectively. The striatal expression suggests that CDNF supports the survival and function of dopaminergic axon terminals. In the midbrain, CDNF might provide a local trophic support for dopaminergic neurons with a paracrine mechanism. Whether CDNF is a specific neurotrophic factor for dopaminergic
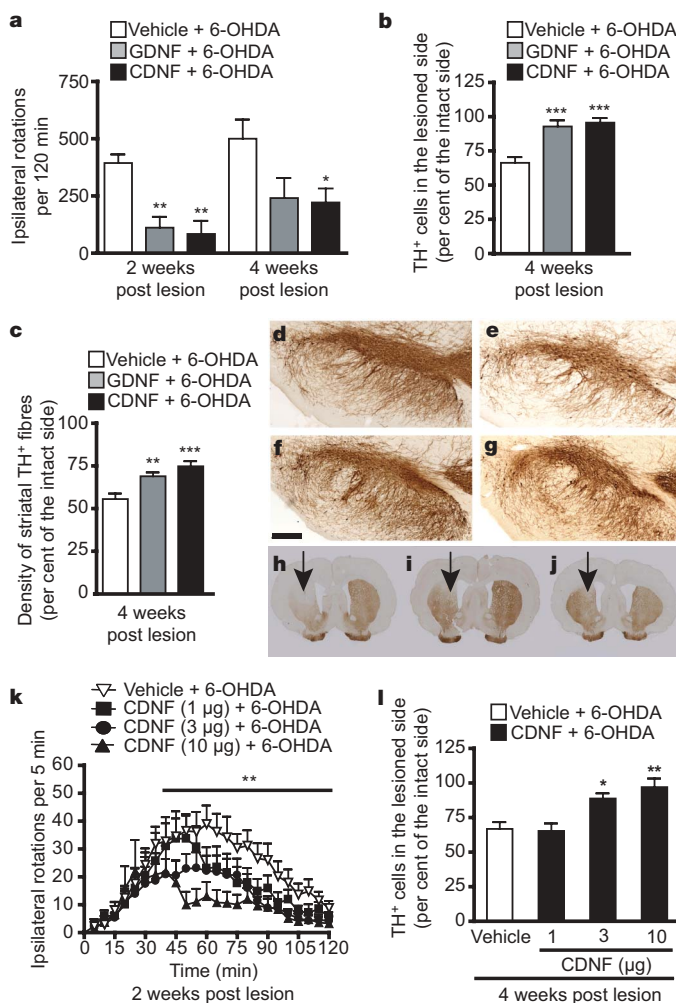

Figure 3 | CDNF protects nigral dopaminergic neurons *in vivo*. The rats were given vehicle, GDNF (10 μg) or CDNF (10 μg) unilaterally into the striatum 6 h before injection of 6-OHDA (8 μg) in to the same location. **a**, Cumulative amphetamine-induced ipsilateral rotations. **b**, Number of tyrosine-hydroxylase (TH)-positive cells in the SNpc. **c**, Density of TH-positive striatal fibres. **d–g**, Photomicrographs of TH-positive cells in the SNpc of an intact rat (**d**), or of a 6-OHDA-lesioned rat pre-treated either with vehicle (**e**), GDNF (**f**) or CDNF (**g**). **h–j**, Photomicrographs of striatal TH-positive fibres in a rat pre-treated either with vehicle (**h**), GDNF (**i**) or CDNF (**j**). The lesioned side is indicated by an arrow. **k**, Dose–response effect of CDNF (1, 3 or 10 μg) on the amphetamine-induced ipsilateral rotations and **l**, on the number of TH-positive cells in SNpc. Scale bar, 300 μm. Mean ± s.e.m.; **a, b, c**, $n = 9$ in each group; **k, l**, $n = 11$ in each group. Tukey/ Kramer post hoc analysis after one-way ANOVA, *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ versus the corresponding control.

neurons or whether it also affects other neuronal populations in the central nervous system will be revealed by future studies. The *in vivo* role of GDNF in the development of dopaminergic neurons has remained unclear[2,16]. Interestingly, the number of dopaminergic neurons of *Gdnf* knockout mice dying at birth does not significantly differ from that of wild type[17–20], suggesting that other neurotrophic factors, including CDNF, the mRNA of which is expressed in embryonic midbrain, have a role in the development of dopaminergic neurons. *Cdnf* mRNA and protein were detected in several adult mouse organs. The role of CDNF in these non-neuronal tissues will be studied elsewhere.

Several neurotrophic factors and growth factors can protect dopaminergic neurons against the toxic effects of 6-OHDA *in vivo*. However, only two neurotrophic factors of the GDNF family, that is,

Figure 4 | CDNF rescues nigral dopaminergic neurons *in vivo*. a, Time course of the experiment. All rats were given a unilateral intrastriatal injection of 6-OHDA (20 µg). Four weeks later, the rats received a single injection of vehicle, GDNF (10 µg) or CDNF (10 µg) to the same location. The rotational behaviour was measured 1 week before and 2, 4, 6 and 8 weeks after the injection of the growth factors. The rats were transcardially perfused after the last rotation experiment. **b,** Cumulative amphetamine-induced ipsilateral rotations. Mean ± s.e.m., $n = 9$ in each group. **c,** Number of TH-positive cells in the SNpc. Mean ± s.e.m., $n = 5$ in each group. *$P < 0.05$ versus the corresponding control.

GDNF[11,21–23] and neurturin[24], are known to possess neurorestorative effects when given after intrastriatal 6-OHDA injection in the rat. Here we show that intrastriatally injected CDNF has both neuroprotective and neurorestorative effects on mature dopaminergic neurons in the rat 6-OHDA model of Parkinson's disease. CDNF treatment before the neurotoxin injection dose-dependently prevented the loss of TH-positive cells in the substantia nigra. In addition, CDNF administered 4 weeks after the 6-OHDA was able to increase the number of TH-positive cells in the substantia nigra by preventing the death of remaining neurons compared with the vehicle-injected controls when measured at 12 weeks post lesion. Striatal injection of 6-OHDA (20 µg) produces a partial lesion of dopaminergic nerve endings which progresses from the injection site retrogradely towards the substantia nigra for 8–16 weeks[25]. The detected restorative effect of CDNF probably reflects the recovery of dopaminergic neurons not fully destroyed by 4 weeks after the 6-OHDA injection. Importantly, CDNF induced behavioural recovery and restored the function of midbrain dopaminergic neurons *in vivo* as efficiently as GDNF. The exact mechanism of CDNF neurorestorative effects will be studied in the future.

All known neurotrophic factors, including GDNF and neurotrophins have effects on embryonic and postnatal peripheral neurons *in vitro*. We did not detect any survival promoting activity of CDNF in primary cultures of P1 mouse superior cervical ganglion sympathetic neurons, E14 rat motoneurons, or E14 and E15 mouse dorsal root ganglion sensory neurons (Supplementary Fig. 6). Our results suggest that CDNF is not a neurite outgrowth and survival promoting factor for neurons in the developing peripheral nervous system, but rather a neurotrophic factor for brain neurons. In this respect, CDNF seems to differ from GDNF, which supports the survival of motoneurons and peripheral sympathetic, parasympathetic, sensory and enteric neurons[2].

Current pharmacotherapy of Parkinson's disease is symptomatic and does not prevent the progressive death of dopaminergic neurons. A future aim is to develop neurorestorative therapies that could halt the degenerative process or even regenerate injured neurons. Because CDNF is able to protect and restore the function of dopaminergic neurons in the rat Parkinson's disease model, it has potential as a therapeutic protein or as a basis for the development of drugs for the treatment of Parkinson's disease.

METHODS

Protein production. Recombinant human CDNF with Flag and $6 \times$ His tags was produced by baculoviral expression in Sf9 insect cells. Secreted CDNF was purified from culture supernatants by nickel affinity chromatography followed by anion exchange chromatography.

Animals. Male Wistar rats (250–280 g; Harlan) were housed under a 12:12 h light:dark cycle at an ambient temperature of 22 °C. Water and rat chow were available *ad libitum*. The experiments were approved by the Committee for Animal Experiments of the University of Helsinki, and by the chief veterinarian of the County Administrative Board (permissions HY 17-03 and HY 14-06).

6-OHDA lesions. All stereotaxic injections of CDNF and GDNF proteins, 6-OHDA and vehicles were done into the left striatum using coordinates relative to the bregma and dura (A/P +1.0, L/M +2.7, D/V -4.0) according to the atlas of Paxinos and Watson[26]. Stereotaxic surgery under isoflurane anaesthesia was performed in the neuroprotection studies in two sessions as described previously[27]. In neuroprotection experiments growth factors were injected 6 h before 6-OHDA. In neurorestoration experiments 6-OHDA was injected first and the growth factors four-weeks later.

Behavioural testing. D-Amphetamine ($2.5\,\text{mg}\,\text{kg}^{-1}$, i.p.) was used to induce rotational activity in rats with unilateral lesion of nigrostriatal dopaminergic nerves. In neuroprotection studies, the rotational behaviour was measured for 2 h at 2 and 4 weeks post lesion as described earlier[9,28]. In neurorestoration studies, behavioural tests were carried out 1 week before (that is, 3 weeks after 6-OHDA injection) and 2, 4, 6 and 8 weeks after the growth factor injections.

Morphological analysis. The number of TH-positive cells in SNpc was estimated using the optical fractionator method[29]. Optical densities of the TH-positive fibres in the striatum were measured from digitalized images using Image Pro-Plus TM software (Version 3.0.1., Media Cybernetics).
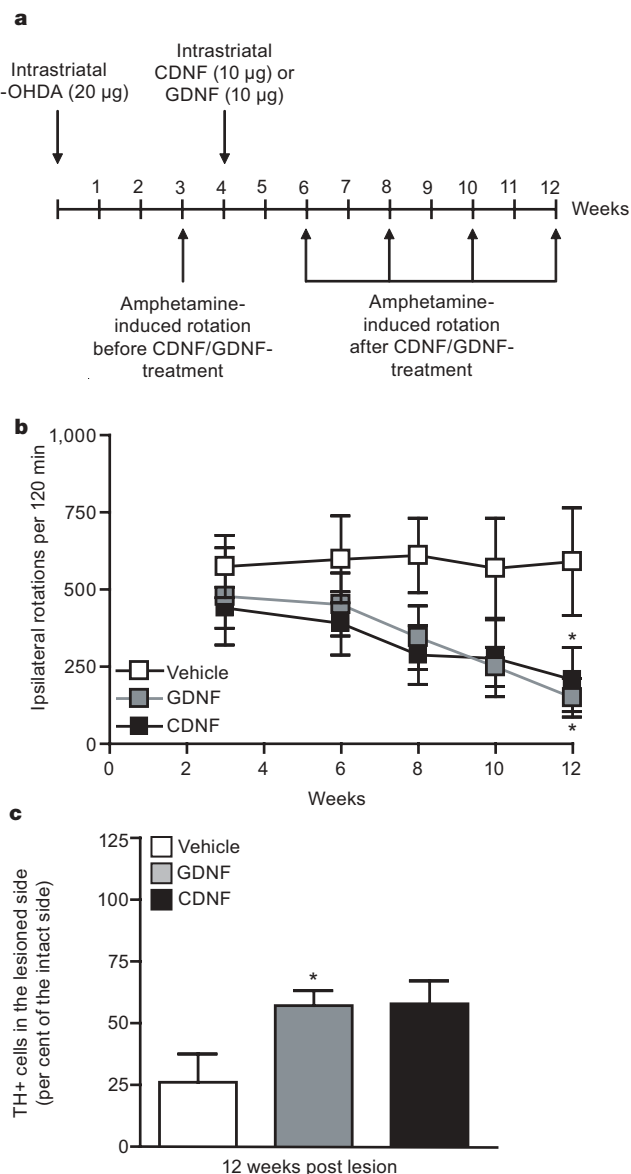
1. Dauer, W. & Przedborski, S. Parkinson's disease: mechanisms and models. *Neuron* **39,** 889–909 (2003).
2. Airaksinen, M. S. & Saarma, M. The GDNF family: signalling, biological functions and therapeutic value. *Nature Rev. Neurosci.* **3,** 383–394 (2002).
3. Huang, E. J. & Reichardt, L. F. Neurotrophins: roles in neuronal development and function. *Annu. Rev. Neurosci.* **24,** 677–736 (2001).
4. Lin, L. F., Doherty, D. H., Lile, J. D., Bektesh, S. & Collins, F. GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. *Science* **260,** 1130–1132 (1993).
5. Petrova, P. S. *et al.* MANF: a new mesencephalic, astrocyte-derived neurotrophic factor with selectivity for dopaminergic neurons. *J. Mol. Neurosci.* **20,** 173–187 (2003).
6. Gill, S. S. *et al.* Direct brain infusion of glial cell line-derived neurotrophic factor in Parkinson disease. *Nature Med.* **9,** 589–595 (2003).
7. Patel, N. K. *et al.* Intraputamenal infusion of glial cell line-derived neurotrophic factor in PD: a two-year outcome study. *Ann. Neurol.* **57,** 298–302 (2005).
8. Lang, A. E. *et al.* Randomized controlled trial of intraputamenal glial cell line-derived neurotrophic factor infusion in Parkinson disease. *Ann. Neurol.* **59,** 459–466 (2006).
9. Ungerstedt, U. & Arbuthnott, G. W. Quantitative recording of rotational behavior in rats after 6-hydroxy-dopamine lesions of the nigrostriatal dopamine system. *Brain Res.* **24,** 485–493 (1970).
10. Kearns, C. M. & Gash, D. M. GDNF protects nigral dopamine neurons against 6-hydroxydopamine *in vivo. Brain Res.* **672,** 104–111 (1995).
11. Aoi, M., Date, I., Tomita, S. & Ohmoto, T. The effect of intrastriatal single injection of GDNF on the nigrostriatal dopaminergic system in hemiparkinsonian rats: behavioral and histological studies using two different dosages. *Neurosci. Res.* **36,** 319–325 (2000).
12. Kirik, D., Rosenblad, C. & Björklund, A. Preservation of a functional nigrostriatal dopamine pathway by GDNF in the intrastriatal 6-OHDA lesion model depends on the site of administration of the trophic factor. *Eur. J. Neurosci.* **12,** 3871–3882 (2000).
13. Åkerud, P., Canals, J. M., Snyder, E. Y. & Arenas, E. Neuroprotection through delivery of glial cell line-derived neurotrophic factor by neural stem cells in a mouse model of Parkinson's disease. *J. Neurosci.* **21,** 8108–8118 (2001).
14. Jaaro, H., Beck, G., Conticello, S. G. & Fainzilber, M. Evolving better brains: a need for neurotrophins? *Trends Neurosci.* **24,** 79–85 (2001).
15. Hidalgo, A., Learte, A. R., McQuilton, P., Pennack, J. & Zhu, B. Neurotrophic and gliatrophic contexts in *Drosophila. Brain Behav. Evol.* **68,** 173–180 (2006).
16. Kramer, E. R. *et al.* Absence of ret signaling in mice causes progressive and late degeneration of the nigrostriatal system. *PLoS Biol.* **5,** e39 (2007).
17. Sánchez, M. P. *et al.* Renal agenesis and the absence of enteric neurons in mice lacking GDNF. *Nature* **382,** 70–73 (1996).
18. Pichel, J. G. *et al.* Defects in enteric innervation and kidney development in mice lacking GDNF. *Nature* **382,** 73–76 (1996).
19. Moore, M. W. *et al.* Renal and neuronal abnormalities in mice lacking GDNF. *Nature* **382,** 76–79 (1996).
20. Granholm, A. C. *et al.* Morphological alterations in the peripheral and central nervous systems of mice lacking glial cell line-derived neurotrophic factor (GDNF): immunohistochemical studies. *J. Neurosci.* **17,** 1168–1178 (1997).
21. Bowenkamp, K. E. *et al.* Glial cell line-derived neurotrophic factor supports survival of injured midbrain dopaminergic neurons. *J. Comp. Neurol.* **355,** 479–489 (1995).
22. Hoffer, B. J. *et al.* Glial cell line-derived neurotrophic factor reverses toxin-induced injury to midbrain dopaminergic neurons *in vivo. Neurosci. Lett.* **182,** 107–111 (1994).
23. Rosenblad, C., Martinez-Serrano, A. & Björklund, A. Intrastriatal glial cell line-derived neurotrophic factor promotes sprouting of spared nigrostriatal dopaminergic afferents and induces recovery of function in a rat model of Parkinson's disease. *Neuroscience* **82,** 129–137 (1998).
24. Rosenblad, C. *et al.* Protection and regeneration of nigral dopaminergic neurons by neurturin or GDNF in a partial lesion model of Parkinson's disease after administration into the striatum or the lateral ventricle. *Eur. J. Neurosci.* **11,** 1554–1566 (1999).
25. Sauer, H. & Oertel, W. H. Progressive degeneration of nigrostriatal dopamine neurons following intrastriatal terminal lesions with 6-hydroxydopamine: a combined retrograde tracing and immunocytochemical study in the rat. *Neuroscience* **59,** 401–415 (1994).
26. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates*. (Academic Press, San Diego, 1997).
27. Kearns, C. M., Cass, W. A., Smoot, K., Kryscio, R. & Gash, D. M. GDNF protection against 6-OHDA: time dependence and requirement for protein synthesis. *J. Neurosci.* **17,** 7111–7118 (1997).
28. Janhunen, S., Tuominen, R. K. & Ahtee, L. Comparison of the effects of nicotine and epibatidine given in combination with nomifensine on rotational behaviour in rats. *Neurosci. Lett.* **381,** 314–319 (2005).
29. West, M. J., Slomianka, L. & Gundersen, H. J. Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *Anat. Rec.* **231,** 482–497 (1991).

Author Contributions P.L., J.L., T.T. and M.S. discovered the new neurotrophic factor. P.L. expressed and purified the protein, studied its tissue expression and together with M.H.V. and M.S. wrote the paper. M.H.V. planned and carried out the neuroprotection and neurorestoration experiments *in vivo.* J.P. raised the CDNF antibodies. V.-M.L. helped in CDNF purification and N.K. did the protein sequencing and mass analysis. J.-O.A. and M.L. did *in vitro* assays with neurons. R.K.T. and S.J. planned and supervised the *in vivo* experiments and M.S. supervised the whole project. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S. (mart.saarma@helsinki.fi).

METHODS

Cloning and RT–PCR analysis. Mouse total RNA was isolated using RNAwiz reagent (Ambion) and human RNAs were obtained from Clontech or were prepared as described[30]. All experiments with human tissues were approved by the local ethical committee. First-strand cDNAs were synthesized with Superscript[II] reverse transcriptase (Invitrogen), according to the manufacturer's instructions, using oligo(dT) primed total RNA (5 µg) or poly(A)$^+$ RNA (1 µg) from different tissues as a template. The following PCR primers were used: m-*Cdnf*-ATG, (5′-ACCATGCGGTGCATCAGTCCAACTGC-3′) and m-*Cdnf*-stop-del, (5′-GAGCTCCGTTTGGGGGTATATC-3′) for mouse *Cdnf*; h-*CDNF*-ATG, (5′-ACCATGTGGTGCGCGAGCCCAGTTGC-3′) and h-*CDNF*-stop-del, (5′-GAGCTCTGTTTTGGGGTGTGTC-3′) or h-*CDNF*-stop, (5′-GATCA-GAGCTCTGTTTTGGGGTGTGTC-3′) for human *CDNF*. PCR reactions were performed using the Expand High Fidelity PCR System and GC-rich PCR System Kit (Roche) for mouse and human *CDNF*, respectively. PCR products were cloned into pCRII (Invitrogen) and sequenced. Human *CDNF* was subcloned into pCR3.1 (without tags) and pcDNA3.1 vector (with carboxy-terminal V5 and 6 × His tags; Invitrogen). To assess the amount of template in the PCR reactions, a control PCR was performed with primers specific for ubiquitously expressed *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) or *PPIA* (peptidylprolyl isomerase A).

Protein production. A cDNA encoding mature human CDNF (amino acids 27–187) was subcloned with amino-terminal Flag and 6 × His tags into pK5039, a derivative of pFASTBAC1 (ref. 31) (Gibco). Recombinant baculoviruses were produced according to the Bac-to-Bac protocol (Invitrogen). Sf9 cells were grown in serum-free SF-900 II medium (Invitrogen) containing antibiotic/antimycotic solution (Gibco) at 28 °C. At 3 days post infection, soluble CDNF was purified from Sf9 culture supernatants by Ni-chromatography. Supernatant was adjusted to pH 7.5 with phosphate-buffered saline (PBS), and incubated with Ni^{2+}-charged Chelating Sepharose Fast Flow (Amersham). After washing with 0.5 M KCl in PBS, proteins were eluted with 0.5 M imidazole, pH 8.8, and concentrated. For further purification with anion exchange chromatography, the sample was diluted with 50 mM Tris-HCl, pH 8.8, and applied on a HiTrap Q HP column (Amersham). CDNF was eluted with a 0 to 1 M NaCl gradient in 50 mM Tris-HCl, pH 8.2. For *in vitro* and *in vivo* assays, the protein buffer was changed to PBS.

Antibody production. CDNF or MANF open reading frame, excluding the signal peptide, was cloned into a T7*lac*-based vector[32,33]. The proteins were purified and used for immunization of two rabbits as previously described[33]. The antiserum was affinity purified by using recombinant CDNF coupled to CNBr Sepharose (Amersham) beads[33].

CDNF immunohistochemistry. Adult male, P10 and P1 NMRI mice brain paraffin sections (8 µm) or cryosections (17 µm) were used. Sections were stained with anti-CDNF antibodies (1:40) using avidin-biotin-peroxidase system (Elite ABC Kit, Vector) and 3,3′-diaminobenzidine (DAB). Adult sections were counterstained with cresyl violet. For double immunofluorescence, mouse anti-NeuN (1:500, MAB377, Chemicon) or mouse anti-tyrosine-hydroxylase (1:2000, MAB318, Chemicon), goat anti-mouse 568 Alexa and donkey anti-rabbit 488 Alexa (1:400, Molecular Probes) antibodies were used with antibodies to CDNF.

6-OHDA lesions. Stereotaxic surgery under isoflurane anaesthesia was performed in the neuroprotection studies in two sessions as described previously[27]. In the first operation, the rats were administered 10 µg of either recombinant human CDNF in 4 µl of PBS or recombinant human GDNF (PeproTech or Amgen) in 4 µl of 10 mM citric acid, pH 5.0, or 4 µl of vehicle (controls; PBS or 10 mM citric acid). In dose-dependence experiments, CDNF was given at doses of 1, 3 or 10 µg in 4 µl of PBS. In the second operation 6 h later, all animals were administered 8 µg per 4 µl of 6-OHDA (Sigma; calculated as a free base, dissolved in ice-cold saline containing 0.02% ascorbic acid). Desipramine (Sigma; 15 mg kg^{-1}, i.p., 1 ml kg^{-1}) was administered before 6-OHDA to prevent the uptake of 6-OHDA into noradrenergic nerve endings, and thus to protect these nerve terminals from destruction. In neurorestoration experiments, the rats received a unilateral intrastriatal injection of 20 µg per 4 µl of 6-OHDA. Four weeks later, the rats were given an intrastriatal injection of CDNF (10 µg in 4 µl of PBS), GDNF (10 µg in 4 µl of 10 mM citric acid) or vehicle.

Tyrosine hydroxylase immunohistochemistry. Rat brains were processed for midbrain and striatal tyrosine-hydroxylase-immunohistochemistry as described previously[34].

SN cell counts. The number of TH-positive cells in SNpc was estimated using the optical fractionator method in combination with the dissector principle and unbiased counting rules[29]. The SNpc was analysed as described previously[35] with Stereo Investigator platform (MicroBrightField) attached to an Olympus BX51 microscope. From each animal, three sections from the central portion of the SNpc, where the medial terminal nucleus was present (level A/P −5.3 mm in the

atlas of Paxinos and Watson[26]), were selected for quantitative analysis. Each reference space was outlined at low power (4×), and cells were counted using a high magnification (60×, oil immersion) objective. Cell numbers were expressed as the mean number per section.

Striatal fibre density. The optical densities of the TH-positive fibres in the striatum were determined from three coronal striatal sections from each rat. Every sixth section between AP +1.6 and AP +0.20 was cut on a freezing microtome and processed for TH-immunohistochemistry. Digitalized images from TH-stained striata were taken with Optronics digital camera and a constant illumination table, and fibre densities were measured using Image Pro-Plus TM software (Version 3.0.1., Media Cybernetics). The optical density analysis was performed under blinded conditions on coded slides. The data are presented as percentage of the intact side, which was defined as 100%.

Statistical analysis. The numbers of ipsilateral rotations and TH-positive cells were analysed using one-way ANOVA followed by Tukey/Kramer's post-hoc test. The numbers of TH-positive cells in the studies of neurorestoration were analysed using a two-tailed Student's *t*-Test. Results from dose–response studies (amphetamine-induced rotation) were analysed using one-way ANOVA for repeated measures followed by Tukey/Kramer's post-hoc test. Results are expressed as mean ± s.e.m. Results were considered significant at $P < 0.05$.

30. Timmusk, T., Belluardo, N., Metsis, M. & Persson, H. Widespread and developmentally regulated expression of neurotrophin-4 mRNA in rat brain and peripheral tissues. *Eur. J. Neurosci.* **5**, 605–613 (1993).
31. Keinänen, K., Jouppila, A. & Kuusinen, A. Characterization of the kainate-binding domain of the glutamate receptor GluR-6 subunit. *Biochem. J.* **330**, 1461–1467 (1998).
32. Peränen, J., Rikkonen, M., Hyvönen, M. & Kääriäinen, L. T7 vectors with a modified T7*lac* promoter for expression of proteins in *Escherichia coli. Anal. Biochem.* **236**, 371–373 (1996).
33. Peränen, J. & Furuhjelm, J. Expression, purification and properties of Rab8 function in actin cortical skeleton organization and polarized transport. *Methods Enzymol.* **329**, 188–196 (2001).
34. Kirik, D., Georgievska, B., Rosenblad, C. & Björklund, A. Delayed infusion of GDNF promotes recovery of motor function in the partial model of Parkinson's disease. *Eur. J. Neurosci.* **13**, 1589–1599 (2001).
35. Sauer, H., Rosenblad, C. & Björklund, A. Glial cell line-derived neurotrophic factor but not transforming growth factor β3 prevents delayed degeneration of nigral dopaminergic neurons following striatal 6-hydroxydopamine lesion. *Proc. Natl Acad. Sci. USA* **92**, 8935–8939 (1995).

LETTERS

TRIC channels are essential for Ca^{2+} handling in intracellular stores

Masayuki Yazawa[1,2], Christopher Ferrante[3], Jue Feng[2], Kazuhiro Mio[4], Toshihiko Ogura[4,5], Miao Zhang[1,2], Pei-Hui Lin[3], Zui Pan[3], Shinji Komazaki[6], Kazuhiro Kato[2], Miyuki Nishi[1,2], Xiaoli Zhao[3], Noah Weisleder[3], Chikara Sato[4], Jianjie Ma[3] & Hiroshi Takeshima[1,2]

Cell signalling requires efficient Ca^{2+} mobilization from intracellular stores through Ca^{2+} release channels, as well as predicted counter-movement of ions across the sarcoplasmic/endoplasmic reticulum membrane to balance the transient negative potential generated by Ca^{2+} release[1–7]. Ca^{2+} release channels were cloned more than 15 years ago[8,9], whereas the molecular identity of putative counter-ion channels remains unknown. Here we report two TRIC (trimeric intracellular cation) channel subtypes that are differentially expressed on intracellular stores in animal cell types. TRIC subtypes contain three proposed transmembrane segments, and form homo-trimers with a bullet-like structure. Electrophysiological measurements with purified TRIC preparations identify a monovalent cation-selective channel. In TRIC-knockout mice suffering embryonic cardiac failure, mutant cardiac myocytes show severe dysfunction in intracellular Ca^{2+} handling. The TRIC-deficient skeletal muscle sarcoplasmic reticulum shows reduced K$^+$ permeability, as well as altered Ca^{2+} 'spark' signalling and voltage-induced Ca^{2+} release. Therefore, TRIC channels are likely to act as counter-ion channels that function in synchronization with Ca^{2+} release from intracellular stores.

In the course of screening membrane proteins participating in cellular Ca^{2+} handling[10,11], we identified a protein with a calculated molecular mass of 33,300 from rabbit skeletal muscle, and named it TRIC type A (TRIC-A; also known as mitsugumin 33). Homology searches in databases revealed an additional structural homologue named TRIC-B. TRIC subtypes show fragmentary sequence identities, as determined by our complementary DNA and *in silico* cloning from various animal species (Supplementary Fig. 1). Northern blotting indicated that TRIC-A is preferentially expressed in excitable tissues, including striated muscle and brain, whereas TRIC-B is present in most mammalian tissues (Fig. 1a). Western blotting of fractionated muscle membranes suggested that TRIC-A is distributed throughout the sarcoplasmic reticulum (SR) but not in the cell-surface membranes (Fig. 1b). In further immunochemical studies, antibodies to TRIC-A decorated the SR and nuclear membranes in skeletal muscle, and TRIC-B behaved as an endoplasmic reticulum (ER)-resident protein in the brain tissues (Supplementary Fig. 1). Therefore, TRIC subtypes are localized to membrane systems associated with intracellular Ca^{2+} stores.

TRIC subtypes show conserved hydropathicity profiles that suggest multiple transmembrane segments (Fig. 1c). In limited proteolysis analysis using membrane vesicles (Supplementary Fig. 2), we found that the amino terminus of TRIC-A is located in the SR/ER lumen, whereas the carboxy terminus of TRIC-A is exposed to the cytoplasm. Further analysis of epitope-tagged recombinant proteins predicted three transmembrane segments in TRIC-A, and also detected the hydrophobic loop as a candidate for an ion-conducting pore between the first and second transmembrane segments. The proposed topology of TRIC subtypes (Fig. 1d) bears an overall resemblance to that of glutamate receptor channels[12].

Affinity chromatography using monoclonal antibodies allowed us to obtain pure TRIC-A protein (>95% purity) from muscle microsomal preparations solubilized with *n*-dodecyl β-D-maltoside (DDM) or digitonin (Supplementary Fig. 3). Treatment of TRIC-A with several chemical crosslinkers generated products with sizes corresponding to dimeric and trimeric assemblies (Fig. 1e and Supplementary Fig. 3). When purified TRIC-A was labelled with colloidal-gold-conjugated Fab fragments, electron microscopy frequently detected antigen–antibody complexes carrying three immunogold particles that protruded at near equal angular intervals (Fig. 1f). Combined computer algorithms, which collect, classify and average negatively stained electron-microscope images to reconstruct a three-dimensional volume[13–16], and the tilt series of the particle images, demonstrated that TRIC-A forms an elongated triangular pyramidal structure (Fig. 1g and Supplementary Figs 4 and 5). Thus, TRIC subtypes form a homo-trimeric structure, as in the cases of the P2X (ref. 17) and bacterial porin channels[18].

Using lipid bilayer reconstitution studies, we observed a functional cation-selective channel with TRIC-A purified from skeletal muscle (Fig. 2a, b). In a recording solution composed of 200 mM KCl (*cis*)/50 mM KCl (*trans*) (see Methods), a slope conductance of 110 ± 14 pS and a reversal potential of -20 ± 1.7 mV were observed, indicating the cation-selective nature of the TRIC-A channel (Fig. 2f). Although the single channel conductance properties of TRIC-A are similar to those of the SR K$^+$ channel reported previously[5,19], other characteristics suggest that the TRIC-A channel might represent a separate channel moiety. First, under bi-ionic conditions of 200 mM KCl (*cis*)/200 mM NaCl (*trans*), outward current was measured at a holding potential of 0 mV, which reverses at a potential of -10 mV (Fig. 2c). This negative reversal potential corresponds to moderate selectivity for K$^+$ over Na$^+$ (permeability ratio of $P_K/P_{Na} = 1.5$), which contrasts to the high K$^+$ selectivity of the SR K$^+$ channel[5,19]. Second, the SR K$^+$ channel is highly sensitive to inhibition by decamethonium[5], whereas the TRIC-A channel is insensitive to this drug at concentrations up to 100 μM (Fig. 2a). Furthermore, application of either CaCl$_2$ or MgCl$_2$ to the recording solutions did not significantly affect the reversal potential or open probability, indicating that the TRIC-A channel is impermeable and insensitive to divalent

[1]Department of Biological Chemistry, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan. [2]Department of Medical Chemistry, Graduate School of Medicine, Tohoku University, Miyagi 980-8575, Japan. [3]Department of Physiology and Biophysics, Robert Wood Johnson Medical School, New Jersey 08854, USA. [4]Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki 305-8568, Japan. [5]PRESTO, Japan Science and Technology Agency, Saitama 332-0012, Japan. [6]Department of Anatomy, Saitama Medical University, Saitama 350-0495, Japan.

cations such as Ca^{2+} and Mg^{2+} (Supplementary Fig. 6). It is likely that TRIC channels either depend on putative interacting partners for certain characteristics, or represent an SR/ER channel that has not been previously reported.

As in the native specimens, the homo-trimeric assembly was also detected in epitope-tagged TRIC-A prepared from bacterial culture (Supplementary Fig. 3). This recombinant TRIC-A also formed a cation channel with kinetic properties indistinguishable from those of the native protein (Fig. 2d, f). Moreover, the application of monoclonal antibody against the C terminus of TRIC-A blocked the channel activity of both recombinant and native TRIC-A preparations (Fig. 2d, e), indicating that the observed channels are unlikely to result from minor impurities in our preparations.

To examine the physiological role of TRIC channels, we generated knockout mice (Supplementary Fig. 7). Mutant mice lacking *Tric-a* were viable and fertile, whereas mutant mice lacking *Tric-b* showed neonatal lethality. By crossing them, double-knockout mice lacking both subtypes (TRIC-DKO) and $Tric-a^{+/-}Tric-b^{+/-}$ (TRIC-DHE)

mice were generated. TRIC-DKO mice exhibited a weak heartbeat at embryonic day (E) 9.5 and subsequent loss of cardiomyocyte viability (Supplementary Fig. 8), leading to terminal embryonic cardiac lethality beyond E10.5 (Fig. 3a). This aggravated lethality indicates that TRIC subtypes share complementary physiological functions. In the looped cardiac tubes from E8.5–9.5 TRIC-DKO embryos, irregular cytoplasmic vacuoles were formed, in particular within the ventricular myocytes (Fig. 3b). Electron microscopy revealed extensively swollen SR/ER structures in TRIC-DKO cardiomyocytes, resulting in the loss of normal rough and smooth ER structures (Fig. 3b and Supplementary Fig. 9). Using fixative solutions supplemented with oxalate to visualize high Ca^{2+} content in intracellular organelles[20], we frequently observed electron-dense Ca-oxalate deposits in the bloated SR/ER in TRIC-DKO myocytes, but not in TRIC-DHE myocytes (Fig. 3c).

Fluorometric Ca^{2+} imaging of TRIC-DKO cardiomyocytes revealed that the amplitudes of spontaneous Ca^{2+} oscillations were remarkably depressed at E8.5, illustrating compromised intracellular Ca^{2+} signalling associated with ablation of TRIC subtypes (Fig. 3d).

Figure 1 | Biochemical characterization of TRIC subtypes. a, Northern analysis of *Tric* subtypes in total RNA preparations (25 μg) from adult mice. **b,** Western analysis in rabbit skeletal muscle membrane preparations (total microsome, T-tubule, light SR, intermediate SR and heavy SR fractions; 2.5 μg) using antibodies to TRIC-A, sarcalumenin (SAR, longitudinal SR marker), ryanodine receptor (RyR, heavy SR marker) and mitsugumin 29 (MG29, T-tubular marker). **c,** Hydropathicity profile of TRIC-A in the Kyte–Doolittle algorithm (window size, 19 residues). M1–M3, putative transmembrane segments. **d,** Topology model of TRIC-A proposed in Supplementary Fig. 2. **e,** Chemical crosslinking of TRIC-A with 1,11-bis-maleimidotetraethyleneglycol. Arrowheads, monomeric, dimeric and trimeric forms. **f,** Electron microscopy images of native TRIC-A bound with gold-conjugated Fab fragment (left panel; scale bar, 200 Å). High-magnification images of the immuno-complexes and their contours are also shown (right panels; scale bar, 100 Å). **g,** Surface representations of three-dimensional reconstruction of native TRIC-A. Scale bar, 50 Å.
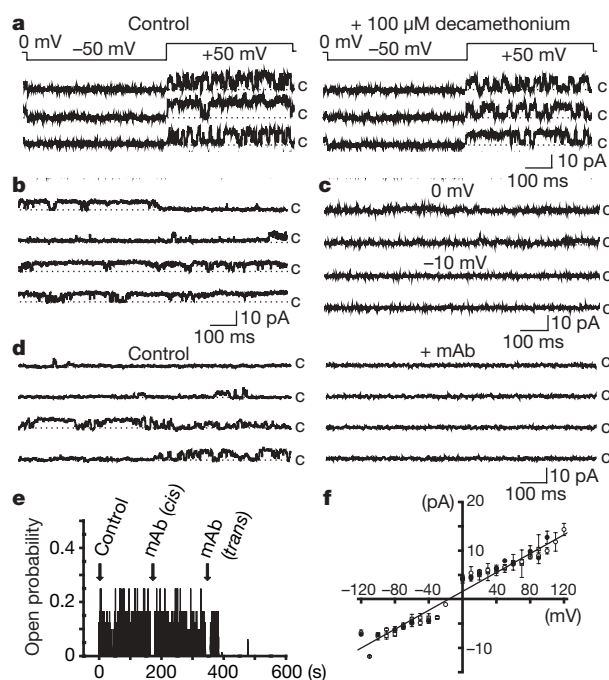

Figure 2 | Electrophysiological characterization of the TRIC-A channel. a, Single-channel recording of native TRIC-A purified using digitonin with 300:50 mM KCl (*cis:trans*) as the current carrier. Representative traces are shown from experiments in the absence (control, $n = 24$) or presence of 100 μM decamethonium ($n = 11$). The closed state (c) is marked with dashed lines. **b,** Current traces of native TRIC-A channel purified with DDM was recorded in a 200:50 mM KCl solution at a holding potential of 0 mV ($n = 12$). **c,** Under bi-ionic conditions of 200 mM KCl (*cis*)/200 mM NaCl (*trans*), TRIC-A current was measured at a holding potential of 0 mV, and became zero at −10 mV ($n = 5$). **d,** Channel activity from recombinant TRIC-A protein was recorded in a 200:50 mM KCl solution at a holding potential of 0 mV (control). In this recording, application of monoclonal antibody MY474 (20 μg ml^{-1}) to the *trans* solution led to complete closure of the channel (+mAb). **e,** Open probability diary plot illustrates that the recombinant TRIC-A channel was sensitive to blockade by MY474 from the *trans* side, as application to the *cis* side has no effect ($n = 3$). In other experiments, addition of the antibody to the *cis* solution could also block the TRIC-A channel ($n = 3$), owing to the reversal of channel orientation in the lipid bilayer. Similar effects of monoclonal antibody were observed in native TRIC-A channels (*cis* effect, $n = 3$, and *trans* effect, $n = 3$). **f,** Current–voltage relationships of native TRIC-A channels purified using digitonin (open circles) or DDM (open squares), and recombinant TRIC-A channels (filled circles) were obtained in a 200:50 mM KCl solution. Solid line represents a slope conductance of 110 ± 14 pS, with a reversal potential of −20.0 ± 1.7 mV (means ± s.e.m.).

Interestingly, TRIC-DKO myocytes displayed significantly larger caffeine-evoked Ca^{2+} transients than the TRIC-DHE myocytes (Fig. 3e). Because caffeine opens ryanodine receptor (RyR) channels and triggers Ca^{2+} release[21,22], the elevated caffeine-evoked Ca^{2+} transients suggest the SR/ER in TRIC-DKO cardiomyocytes is overloaded with Ca^{2+}, consistent with the Ca-oxalate deposits observed by electron microscopy. In Ca^{2+}-induced Ca^{2+} release (CICR) during cardiac excitation–contraction coupling, Ca^{2+} influx through dihydropyridine-sensitive channels (DHPR) activates the cardiac subtype of RyR (RyR2)[21–23]. In embryonic cardiomyocytes bearing immature intracellular stores, spontaneous Ca^{2+} oscillations are primarily produced by Ca^{2+} influx, with RyR2-mediated CICR acting to amplify Ca^{2+} signalling[20,24]. Both TRIC-DKO and RyR2-knockout mice display similar cardiac arrest phenotypes and share the swollen and overloaded SR/ER in embryonic cardiomyocytes[20]. These nearly identical pathologies, coupled with the observation that the E8.5 TRIC-DKO hearts retained normal expression of major Ca^{2+} store-related proteins, including RyR2, DHPR, SR/ER Ca^{2+}-ATPase (SERCA) and Na^+/Ca^{2+}-exchanger (Supplementary Fig. 8), suggest that depressed RyR2-mediated CICR weakens spontaneous oscillations and probably generates SR Ca^{2+} overloading in TRIC-DKO cardiomyocytes.

To examine the contribution of TRIC channels to Ca^{2+} signalling in skeletal muscle, we used osmotic stress to induce Ca^{2+} sparks, the elemental events of SR Ca^{2+} release, in isolated adult skeletal muscle fibres[25,26]. Line-scan imaging (Fig. 4a, b) detected a significant difference in Ca^{2+} spark amplitude between wild-type (WT) mice ($\Delta F/F_0 = 1.06 \pm 0.01$) and $Tric\text{-}a^{-/-}\text{-}Tric\text{-}b^{+/-}$ (mutant) mice (0.66 ± 0.01). In addition, the full duration at half maximum of Ca^{2+} sparks increased from 230 ± 17 ms in WT muscle to 301 ± 17 ms in mutant muscle ($P < 0.005$), probably reflecting a reduced Ca^{2+}-dependent inactivation mechanism of skeletal muscle RyR (RyR1)[3,21,22] associated with the impaired Ca^{2+} spark signalling. Using the silicone-grease embedding method to measure voltage-induced Ca^{2+} release (VICR) in muscle fibres[3,21,27], we examined the kinetic properties of SR Ca^{2+} release invoked by depolarization pulses to -20 mV from a holding potential of -80 mV. The kinetics of VICR was significantly slower in mutant muscle compared to WT

muscle (Fig. 4c, d), indicating that TRIC deficiency leads to compromised SR Ca^{2+} release in skeletal muscle. We also found that the amplitudes of VICR were similar between WT ($\Delta F/F_0 = 2.48 \pm 0.08$) and mutant ($2.58 \pm 0.14$, $P = 0.52$) muscles, which would be expected because the amplitude of VICR at submaximal stimulation (-20 mV test pulse) is dependent on the recruitment of multiple RyRs to produce global Ca^{2+} release, whereas the effects of counter-current activity on Ca^{2+} release can be more effectively resolved at the level of individual Ca^{2+} sparks (Fig. 4a).

To test directly the contribution of TRIC channels to ionic permeability across the SR, we monitored the changes in membrane potential using a fluorescent voltage indicator, di-8-ANEPPQ. In SR vesicles prepared from hindlimb muscle and Ca^{2+}-loaded under ATP-containing conditions, the fluorescent dye detected a slightly negative potential of the luminal side, which was enhanced by thapsigargin-induced Ca^{2+} leak and counteracted by the K^+ ionophore, valinomycin (Fig. 4e, f). Membrane potential was significantly elevated under both resting/Ca^{2+}-loaded and thapsigargin-induced leak conditions in SR vesicles prepared from $Tric\text{-}a^{-/-}$ mice (Fig. 4f). This elevated potential was largely abrogated by valinomycin (Fig. 4g). Moreover, A23187, a Ca^{2+} ionophore, could return the elevated membrane potential in $Tric\text{-}a^{-/-}$ vesicles to similar levels observed in WT vesicles (data not shown). Taken together, TRIC-A deficiency seems to reduce K^+ permeability without affecting the basal Ca^{2+} permeability across the SR membrane, consistent with the monovalent-cation-selective nature of the TRIC-A channel (Fig. 2).

TRIC channels seem to predominantly be permeated by K^+ under intracellular conditions. Results from our mutant mice indicate that TRIC channels support RyR2-mediated CICR in cardiomyocytes and RyR1-mediated VICR in skeletal muscle, by providing certain aspects of the physiological mechanism for counter-ion movement during rapid SR/ER Ca^{2+} release (Fig. 4h). In addition to TRIC channels, other molecular components may also contribute to this mechanism, as caffeine-evoked Ca^{2+} release remains active in TRIC-DKO cardiomyocytes and VICR is only partially compromised in TRIC-deficient skeletal muscle. For example, as a cation-selective channel
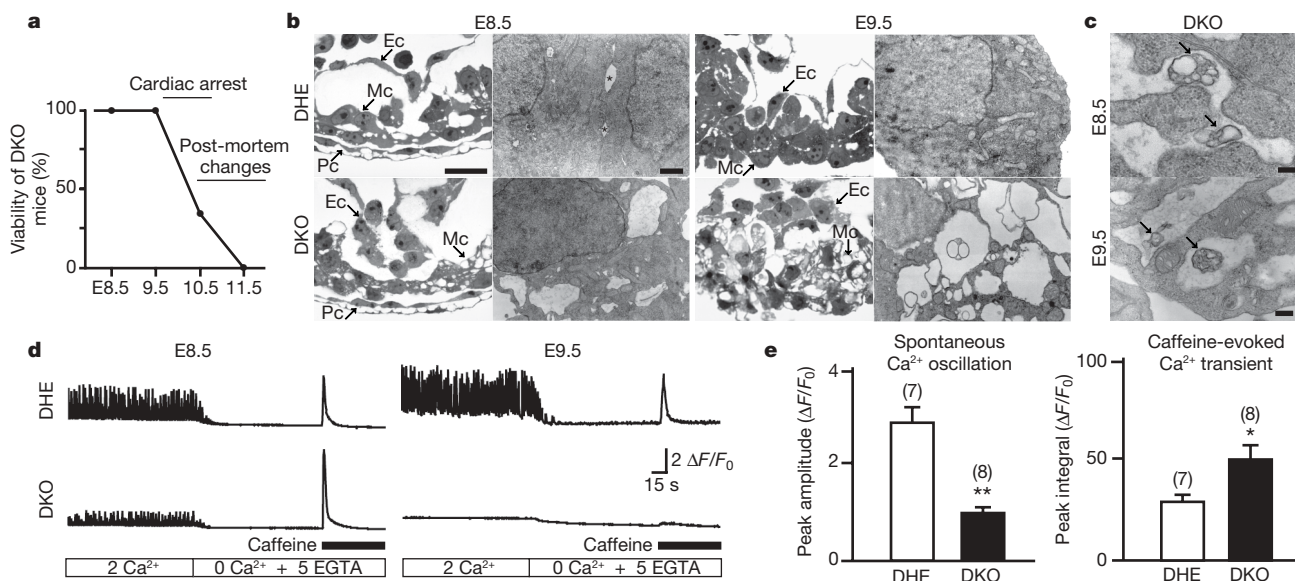

Figure 3 | Physiological abnormalities in TRIC-knockout hearts. a, TRIC-DKO embryos exhibited weak heartbeats at E9.5, stopped beating at ~E10.5, and post-mortem autolysis at ~E11.5. **b,** Histological and ultrastructural abnormalities in TRIC-DKO embryonic hearts. Pc, pericardium; Mc, myocardium; Ec, endocardium; asterisks, intercellular space. Representative data from at least three embryos are shown. Scale bars, 20 μm (left panels) and 1 μm (right panels). **c,** Formation of Ca-oxalate deposits (arrows) in the swollen SR/ER of TRIC-DKO cardiomyocytes. No such precipitates were

observed in control embryos. Scale bar, 0.2 μm. **d,** Fluo-4 Ca^{2+} imaging of TRIC-DKO cardiomyocytes in bathing solutions containing 2 mM $CaCl_2$, 5 mM EGTA or 20 mM caffeine. Representative data from at least seven embryos are shown. **e,** Weak spontaneous Ca^{2+} oscillations and enhanced caffeine-evoked Ca^{2+} transients in E8.5 TRIC-DKO cardiomyocytes. Data represent the mean ± s.e.m., and *n* values are shown in parentheses. *$P < 0.05$ and **$P < 0.01$ in *t*-test.
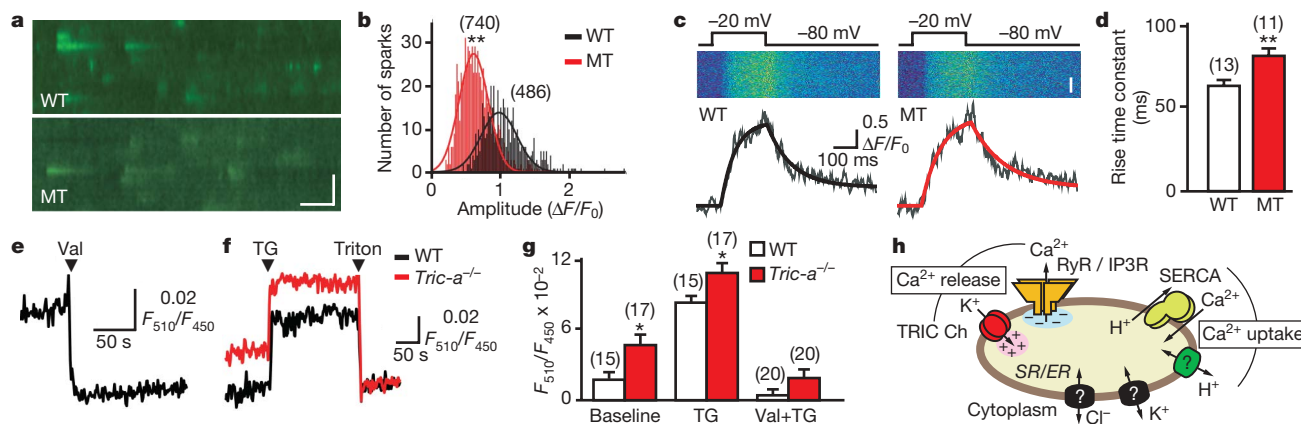
Figure 4 | Compromised Ca²⁺ release and altered membrane potential in TRIC-deficient skeletal muscle SR. a, Ca^{2+} spark images from wild type (WT) and $Tric\text{-}a^{-/-}Tric\text{-}b^{+/-}$ mutant (MT) skeletal muscle fibres. Scale bars, 10 μm and 2 s. **b**, Amplitude histogram of Ca^{2+} sparks were fitted with a gaussian distribution function. **c**, Measurement of VICR in WT and MT fibres under voltage-clamp conditions. Scale bar, 3 μm. Fluorescence traces were fitted by two exponential functions (black and red lines) with rise time constants of 55.3 ms for WT and 76.8 ms for MT. **d**, The rise time constant for VICR in MT fibres (81.9 ± 4.5 ms) was significantly elevated above WT controls (63.9 ± 3.0 ms). **e**, Fluorometric measurements of SR membrane potential using di-8-ANEPPQ. Decrease in the fluorescence ratio following addition of valinomycin (Val, 1 μM) reflects K^+ influx into WT SR vesicles. **f**, Inhibition of SERCA activity by thapsigargin (TG, 2 μM) induced Ca^{2+} leakage and enhanced negative potential across SR vesicles. Data were normalized to levels after the addition of Triton X-100 (0.2%). **g**, Summary data for di-8-ANEPPQ fluorescence at initial levels under Ca^{2+}-loaded conditions (Baseline) and following TG application and Val+TG application. The data in **b**, **d** and **g** are presented as mean ± s.e.m. and n values are shown in parentheses. *$P < 0.05$ and **$P < 0.01$ in t-test or ANOVA. **h**, Proposed role of TRIC channels in Ca^{2+} handling of intracellular stores. Physiological Ca^{2+} release mediated by RyR and IP3-receptor (IP3R) channels probably requires counter-ion movement to neutralize the accumulation of transient negative potential across the SR/ER. Compromised Ca^{2+} spark signalling and VICR in TRIC-deficient muscle indicate that permeation of K^+ or Na^+ through TRIC channels probably contributes to this mechanism. Several H^+, K^+ or Cl^--selective channels were proposed on the ER/SR from electrophysiological and biochemical data, but their molecular identities and biological functions remain to be elucidated.

conducting both Ca^{2+} and monovalent cations[3,22], RyR channels may contribute to both Ca^{2+} release and counter-ion permeation while locked in an open state. SERCA is responsible for Ca^{2+} uptake and catalyses Ca^{2+} and H^+ counter-transport[28], which is probably supported by the high H^+ permeability across the SR/ER membrane[29]. In addition, several K^+ and Cl^--selective channels on the SR/ER that were previously identified by other investigators[5,30] could also participate in neutralizing the transient negative membrane potential generated by Ca^{2+} release. Ubiquitously expressed TRIC channels seem essential for physiological Ca^{2+} release that regulates important cellular functions, and therefore could be unique targets in pharmaceutical or agrichemical applications.

METHODS SUMMARY

TRIC channel purification. Microsomal vesicles from rabbit skeletal muscle were solubilized with digitonin or DDM, and native TRIC-A protein was purified using an affinity column conjugated with monoclonal antibody MY474. Recombinant TRIC-A carrying both 6× histidine and haemagglutinin (HA) tags at the C terminus was purified from *Escherichia coli* through a combination of Ni^{2+}- and anti-HA affinity columns.

Electron microscopy analysis. Purified TRIC-A was decorated with immunogold labelled Fab fragments derived from antibody MY474 and analysed using electron microscopy. The three-dimensional image reconstruction of TRIC-A from negatively stained electron microscopy particle images was conducted by our established SPINNS method[13-15].

Lipid bilayer reconstitution. Electrophysiological characterization of TRIC-A preparations was conducted using conventional methods for lipid bilayer recordings[5,19]. Purified TRIC-A protein was added to the *cis* side of a lipid bilayer membrane formed with 50% phosphatidylserine and 50% phosphatidylethanolamine.

Knockout mouse production. TRIC-A and TRIC-B knockout mice were generated by our standard methods. TRIC-DKO embryos generated were subjected to anatomical, biochemical and Ca^{2+} imaging analyses[11,20].

Muscle Ca²⁺ measurements. Flexor digitorum brevis (FDB) fibres were isolated from adult mice, and Ca^{2+} sparks and VICR were monitored[25-27]. Muscle fibres were loaded with Fluo-4 AM, and Ca^{2+} signalling evoked by osmotic shock or depolarization was captured by line-scan imaging with a confocal microscope.

SR membrane potential measurements. SR vesicles prepared from mouse hindlimb muscle were resuspended in a Cl^--free solution supplemented with

0.5 μM di-8-ANEPPQ (Invitrogen). Potential-dependent fluorescence changes in response to valinomycin, thapsigargin, A23187 or TritonX-100 were monitored at an emission wavelength of 570 nm with dual excitation wavelengths of 450 and 510 nm. Quantitative changes in membrane potential were assayed by changes in the ratio of fluorescence intensity (F_{510}/F_{450}).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Ebashi, S. Excitation-contraction coupling. *Annu. Rev. Physiol.* **38**, 293–313 (1976).
2. Berridge, M. J. The endoplasmic reticulum: a multifunctional signaling organelle. *Cell Calcium* **32**, 235–249 (2002).
3. Fill, M. & Copello, J. A. Ryanodine receptor calcium release channels. *Physiol. Rev.* **82**, 893–922 (2002).
4. Somlyo, A. V., Shuman, H. & Somlyo, A. P. Composition of sarcoplasmic reticulum *in situ* by electron probe X-ray microanalysis. *Nature* **268**, 556–558 (1977).
5. Coronado, R. & Miller, C. Decamethonium and hexamethonium block K^+ channels of sarcoplasmic reticulum. *Nature* **288**, 495–497 (1980).
6. Fink, R. H. & Veigel, C. Calcium uptake and release modulated by counter-ion conductances in the sarcoplasmic reticulum of skeletal muscle. *Acta Physiol. Scand.* **156**, 387–396 (1996).
7. Meissner, G. Monovalent ion and calcium ion fluxes in sarcoplasmic reticulum. *Mol. Cell. Biochem.* **55**, 65–82 (1983).
8. Takeshima, H. *et al.* Primary structure and expression from complementary DNA of skeletal muscle ryanodine receptor. *Nature* **339**, 439–445 (1989).
9. Furuichi, T. *et al.* Primary structure and functional expression of the inositol 1,4,5-trisphosphate-binding protein P400. *Nature* **342**, 32–38 (1989).
10. Takeshima, H. *et al.* Mitsugumin29, a novel synaptophysin family member from the triad junction in skeletal muscle. *Biochem. J.* **331**, 317–322 (1998).
11. Takeshima, H. *et al.* Junctophilins: a novel family of junctional membrane complex proteins. *Mol. Cell* **6**, 11–22 (2000).
12. Dani, J. A. & Mayer, M. L. Structure and function of glutamate and nicotinic acetylcholine receptors. *Curr. Opin. Neurobiol.* **5**, 310–317 (1995).
13. Ogura, T. & Sato, C. Auto-accumulation method using simulated annealing enables fully automatic particle pickup completely free from a matching template or learning data. *J. Struct. Biol.* **146**, 344–358 (2004).
14. Ogura, T. & Sato, C. Automatic particle pickup method using a neural network has high accuracy by applying an initial weight derived from eigenimages: a new reference free method for single-particle analysis. *J. Struct. Biol.* **145**, 63–75 (2004).

15. Ogura, T., Iwasaki, K. & Sato, C. Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking. *J. Struct. Biol.* **143**, 185–200 (2003).

16. van Heel, M. *et al.* Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.* **33**, 307–369 (2000).

17. Mio, K. *et al.* Visualization of the trimeric P2X$_2$ receptor with a crown-capped extracellular domain. *Biochem. Biophys. Res. Commun.* **337**, 998–1005 (2005).

18. Cowan, S. W. *et al.* Crystal structures explain functional properties of two *E. coli* porins. *Nature* **358**, 727–733 (1992).

19. Liu, Q.-Y. *et al.* Reconstitution of the solubilized cardiac sarcoplasmic reticulum potassium channel identification of a putative M_r ~80 kDa polypeptide constituent. *FEBS Lett.* **291**, 13–16 (1991).

20. Takeshima, H. *et al.* Embryonic lethality and abnormal cardiac myocytes in mice lacking ryanodine receptor type 2. *EMBO J.* **17**, 3309–3316 (1998).

21. Endo, M. Calcium release from the sarcoplasmic reticulum. *Physiol. Rev.* **57**, 71–108 (1977).

22. Meissner, G. Ryanodine receptor/Ca^{2+} release channels and their regulation by endogenous effectors. *Annu. Rev. Physiol.* **56**, 485–508 (1994).

23. Fabiato, A. Calcium-induced release of calcium from the cardiac sarcoplasmic reticulum. *Am. J. Physiol.* **245**, C1–C14 (1983).

24. Antoon, F. M. *et al.* Presence of functional sarcoplasmic reticulum in the developing heart and its confinement to chamber myocardium. *Dev. Biol.* **223**, 279–290 (2000).

25. Wang, X. *et al.* Uncontrolled calcium sparks act as a dystrophic signal for mammalian skeletal muscle. *Nature Cell Biol.* **7**, 525–530 (2005).

26. Weisleder, N. *et al.* Muscle aging is associated with compromised Ca^{2+} spark signaling and segregated intracellular Ca^{2+} release. *J. Cell Biol.* **174**, 639–645 (2006).

27. Collet, C. *et al.* Intracellular calcium signals measured with indo-1 in isolated skeletal muscle fibers from control and *mdx* mice. *J. Physiol. (Lond.)* **520**, 417–429 (1999).

28. Toyoshima, C. & Inesi, G. Structural basis of ion pumping by Ca^{2+}-ATPase of the sarcoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 269–292 (2004).

29. Meissner, G. & Young, R. C. Proton permeability of sarcoplasmic reticulum vesicles. *J. Biol. Chem.* **255**, 6814–6819 (1980).

30. Kourie, J. I. *et al.* Characteristics of two types of chloride channel in sarcoplasmic reticulum vesicles from rabbit skeletal muscle. *Biophys. J.* **70**, 202–221 (1996).

Author Contributions M.Y., J.F. and M.Z. conducted biochemical experiments and characterized TRIC-DKO mice. K.M., T.O. and C.S. reconstructed the three-dimensional structure. M.Y., Z.P. and J.M. conducted bilayer measurements. C.F., P-H.L., N.W., X.Z. and J.M. characterized TRIC-deficient skeletal muscle. S.K. was responsible for histology. K.K., M.N. and H.T. identified TRIC subtypes and produced knockout mice. H.T. oversaw the project.

METHODS

TRIC channel purification. For purification of native TRIC-A, total microsomes were prepared from rabbit skeletal muscle[31] and solubilized in a buffer containing 1% DDM, 0.25 M sucrose, 2 mM 2-mercaptoethanol (2ME), 50 mM Tris-HCl (pH 7.4) and protease inhibitors. After the addition of NaCl (0.15 M in final concentration), insoluble materials were removed by ultracentrifugation and solubilized proteins were applied to Protein G Sepharose (Amersham Biosciences) crosslinking with monoclonal antibody MY474. After washing the resin with a buffer containing 0.1% DDM, 0.25 M sucrose, 2 mM 2ME, 0.3 M NaCl, 50 mM Tris-HCl (pH 7.4) and protease inhibitors, TRIC-A was eluted with the synthetic peptide (100 μM), which is composed of the rabbit TRIC-A C-terminal 19 residues carrying the antigenic epitope of antibody MY474. The peptide was removed by gel filtration (Superdex 200, Amersham Biosciences) with a running buffer containing 0.05% DDM, 0.25 M sucrose, 0.15 M NaCl, 20 mM NaPi (pH 7.0) and protease inhibitors. By using similar protocols, native TRIC-A was also prepared with digitonin as a detergent in place of DDM. These preparations were used in crosslinking experiments, lipid bilayer measurements and electron microscopy (EM) imaging analysis.

For production of recombinant protein, the cDNA fragment encoding rabbit TRIC-A carrying the C-terminal haemagglutinin (HA)-tag was subcloned into the pQE60 vector (Qiagen) for histidine-tagged protein expression. Cultured bacteria were treated with lysozyme and sonicated, and recombinant TRIC-A was then solubilized with 1% digitonin. After the removal of cellular debris, the lysate was applied to Ni-Sepharose (Amersham Biosciences), and recombinant protein was eluted with a buffer containing 0.1% digitonin, 20 mM NaPi (pH 7.4), 0.5 M NaCl, 500 mM imidazole, 1 mM EDTA, 2 mM 2ME and proteinase inhibitors. The eluted solution was diluted to reduce the salt concentration and applied to anti-HA agarose (Sigma). After washing the resin, recombinant TRIC-A was recovered using a buffer containing 0.1% digitonin, 20 mM NaPi (pH 7.4), 0.5 M NaCl, 0.1 mM EDTA, 0.1 mg ml^{-1} HA peptide (Sigma) and proteinase inhibitors. The peptide was removed by the Ni-Sepharose affinity chromatography and dialysis, and purified preparations were subjected to crosslinking and lipid bilayer experiments.

Membrane topology assay. Membrane vesicles were prepared from mouse skeletal muscle[32] and HEK293 cells[33] under isotonic conditions, and the transmembrane topology of the TRIC-A channel was examined as essentially described previously[34]. Membrane vesicles were suspended in the buffer containing 10 mM Hepes-KOH (pH 7.4), 100 mM NaCl, 10 mM KCl, 1.5 mM MgCl$_2$, 5 mM sodium EDTA, 5 mM sodium EGTA and 250 mM sucrose, and treated with varying concentrations of proteinase K in the absence or presence of 1% Triton X-100. The reactions were stopped by adding phenylmethylsulphonyl fluoride and the degradability of TRIC-A proteins was analysed by immunoblotting.

EM image analysis. Fab fragment was prepared from antibody MY474 using a commercially available kit (Immunopure Fab preparation kit, Pierce Biotechnology), reacted with colloidal gold particles (British BioCell International) and purified by gel filtration[35]. For the immuno-decoration, purified TRIC-A with DDM was reacted with the Fab fragment at a molecular ratio of 1:5. TRIC-A preparations with or without the immuno-decoration were then applied to thin carbon films, negatively stained with 2% uranyl acetate solution and analysed using an electron microscope (JEM-100CX, JEOL) at 55,100× magnification with a 100-kV acceleration voltage. Images were recorded on SO-163 films (Eastman Kodak) and digitized with a Scitex Leafscan 45 scanner (Leaf systems) at a pixel size of 1.82 Å at the specimen level.

Three-dimensional-structure reconstruction. We developed a single particle image analysis method using neural networks[14,15] and simulated annealing[13], namely SPINNS. For 3D reconstruction of the TRIC-A channel, image analysis was performed using combined SPINNS and IMAGIC V algorithms[16]. The TRIC-A projections were primarily selected in 100 × 100 pixel subframes using the auto-accumulation method with simulated annealing[13]. Selected 330 particles were rotated at 10° increments to generate 330 × 36 images, and they were used to train the three-layer pyramidal-type neural network[14]. The trained neural network further selected 10,998 particles and the obtained images were aligned rotationally and translationally[16,36] using the reference free method[14]. Because of the limited resolution, the contrast transfer function (CTF) correction was not performed in this analysis. The images were circularly masked with a diameter of 80 pixels to exclude the perimeter and sorted into 250 classes by the modified growing neural gas network method, which enables accurate and robust classification against noise because of its flexible neural network structure[15]. The resulting averages were used as new references and the cycle from alignment to classification was repeated 14 times. Next, the orientational Euler angles of the class averages were determined from the sinograms using cross-correlation functions[16] assuming a three-fold symmetry. They were used to calculate an initial 3D structure by the SIRT method[36]. The 3D map was further refined by the

projection matching method for 10 cycles[36]. The FSC was used to assess the resolution of the final 3D map[37] at the threshold of 0.5.

Lipid bilayer measurements. Channel measurements were conducted as described previously[38]. Purified TRIC-A preparations (~0.2 μg) were added to the *cis* side of a lipid bilayer membrane formed across a 200-μm-diameter aperture. The lipid bilayer-forming solution contained 50% phosphatidylserine and 50% phosphatidylethanolamine (Avanti Polar Lipids) at a concentration of 40 mg ml^{-1} in *n*-decane (Sigma). Recording solutions were buffered at pH 7.4 with 10 mM HEPES-Tris. Channel recordings were made with Axopatch 200A patch-clamp unit (Axon Instruments), and the data were digitized at 1 kHz and filtered at 1 kHz (low-pass) using a 16 bit A/D-D/A converter and analysed with pClamp software (Axon Instruments).

Knockout mouse production. Knockout mice were generated using our standard methods[39] using J1 embryonic stem cells[40]. TRIC-DKO embryos generated were subjected to anatomical, biochemical and Ca^{2+} imaging analyses[11,20]. PCR primers used for mouse genotyping (Supplementary Fig. 7) were A-1 (TCAGTGGCGAGGGAGCATTCGTCG), A-2 (GTCCTCCTAACCGTAAAC-AAGAGC), B-1 (GTCATGGAGTACCCGTGGGATGATC) and B-2 (CCCTCT-CCTGGTTCAACACGCTAGC).

Muscle Ca^{2+} measurements. Flexor digitorum brevis (FDB) fibres were isolated from adult mice, and Ca^{2+} sparks and VICR were monitored[25–27]. Muscle fibres were loaded with Fluo-4 AM, and Ca^{2+} signalling evoked by osmotic shock or depolarization was captured by line-scan imaging with a confocal microscope. For voltage clamp, major portions of both ends of the fibre were coated with silicone grease, and Ca^{2+} transients were elicited by repeated 200-ms test pulses to –20 mV from a holding potential of –80 mV.

SR membrane potential measurements. SR vesicles were prepared from 3–6-month-old hindlimb muscle from either wild-type ($n = 20$) or *Tric-a$^{-/-}$* ($n = 13$) mice as described previously[31]. SR vesicles (60 μg protein) were resuspended in 2 ml Cl$^-$-free SR-loading buffer (in mM: K glutamate 107.8, EGTA-KOH 2, MgSO$_4$ 6.6, ATP 5.4, creatine phosphate 15, Ca gluconate 0.98, BES-KOH 20, pH 7.2, free Ca^{2+} level of pCa = 6.6), supplemented with 0.5 μM di-8-ANEPPQ (Invitrogen) for 6 min at room temperature with constant stirring. Potential-dependent fluorescence changes in response to 1 μM valinomycin, 2 μM thapsigargin, 40 μM A23187 or 0.2% TritonX-100 were monitored in a DeltaRAM fluorometer (Photon Technology International) with an emission wavelength of 570 nm and a ratio of the excitation intensity of F_{510}/F_{450}.

31. Saito, A. *et al.* Preparation and morphology of sarcoplasmic reticulum terminal cisternae from rabbit skeletal muscle. *J. Cell Biol.* **99**, 875–885 (1984).
32. Fernandez, J. L., Rosemblatt, M. & Hidalgo, C. Highly purified sarcoplasmic reticulum vesicles are devoid of Ca^{2+}-independent ('basal') ATPase activity. *Biochim. Biophys. Acta* **599**, 552–568 (1980).
33. Zhang, M. *et al.* Calumin, a novel Ca^{2+}-binding transmembrane protein on the endoplasmic reticulum. *Cell Calcium* (in the press)..
34. Feramisco, J. D., Goldstein, J. L. & Brown, M. S. Membrane topology of human Insig-1, a protein regulator of lipid synthesis. *J. Biol. Chem.* **279**, 8487–8496 (2004).
35. Faulk, W. P. & Taylor, G. M. An immunocolloid method for the electron microscope. *Immunochemistry* **8**, 1081–1083 (1971).
36. Frank, J. *Three-dimensional Electron Microscopy of Macromolecular Assemblies* (Oxford Univ. Press, New York, 2006).
37. Harauz, G. & van Heel, M. Exact filters for general geometry three dimensional reconstruction. *Optik* **73**, 146–156 (1986).
38. Ma, J. Block by ruthenium red of the ryanodine-activated calcium release channel of skeletal muscle. *J. Gen. Physiol.* **102**, 1031–1056 (1993).
39. Takeshima, H. *et al.* Excitation-contraction uncoupling and muscular degeneration in mice lacking functional skeletal muscle ryanodine-receptor gene. *Nature* **369**, 556–559 (1994).
40. Li, E., Bestor, T. H. & Jaenish, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).

Intronic microRNA precursors that bypass Drosha processing

J. Graham Ruby[1,2]*, Calvin H. Jan[1,2]* & David P. Bartel[1,2]

MicroRNAs (miRNAs) are ~22-nucleotide endogenous RNAs that often repress the expression of complementary messenger RNAs[1]. In animals, miRNAs derive from characteristic hairpins in primary transcripts through two sequential RNase III-mediated cleavages; Drosha cleaves near the base of the stem to liberate a ~60-nucleotide pre-miRNA hairpin, then Dicer cleaves near the loop to generate a miRNA:miRNA* duplex[2,3]. From that duplex, the mature miRNA is incorporated into the silencing complex. Here we identify an alternative pathway for miRNA biogenesis, in which certain debranched introns mimic the structural features of pre-miRNAs to enter the miRNA-processing pathway without Drosha-mediated cleavage. We call these pre-miRNAs/introns 'mirtrons', and have identified 14 mirtrons in *Drosophila melanogaster* and another four in *Caenorhabditis elegans* (including the reclassification of *mir-62*). Some of these have been selectively maintained during evolution with patterns of sequence conservation suggesting important regulatory functions in the animal. The abundance of introns comparable in size to pre-miRNAs appears to have created a context favourable for the emergence of mirtrons in flies and nematodes. This suggests that other lineages with many similarly sized introns probably also have mirtrons, and that the mirtron pathway could have provided an early avenue for the emergence of miRNAs before the advent of Drosha.

While examining sequencing data of small RNAs from *D. melanogaster*[4], we observed clusters of small RNAs originating from the outer edges of an annotated 56-nucleotide (56-nt) intron (Fig. 1a). These sets of reads (each read representing an independently sequenced complementary DNA) had properties similar to those observed previously for miRNA:miRNA* duplexes[5], in that each set had a more consistent 5′ than 3′ terminus, and the two sets were complementary to each other, with the dominantly abundant species of each set forming 2-nt 3′ overhangs when paired to each other. Moreover, the sequence and predicted secondary structure of the intron were conserved in a pattern resembling that of pre-miRNAs[6] (Fig. 1b, c). We annotated this locus as *mir-1003*.

Despite these clearly miRNA-like properties, semblance to canonical miRNA primary transcripts (pri-miRNAs) stopped abruptly at the borders of the intron. Pairing at the base of the hairpin did not extend beyond the miRNA:miRNA* duplex—that is, beyond the splice sites. In place of extended pairing, which is needed for pri-miRNA cleavage by Drosha (ref. 7), the intron had conserved canonical splice sites (Fig. 1b), leading to the model that this miRNA did not arise from a canonical miRNA biogenesis pathway but instead arose from an alternative pathway in which splicing, rather than Drosha, defined the pre-miRNA (Fig. 1d). Consistent with this model, spliced lariats linearized by the lariat debranching enzyme bear 5′ monophosphates[8] and 3′ hydroxyls[9], the same moieties found in pre-miRNAs[1,3,10].

Thirteen additional pre-miRNAs/introns, termed mirtrons, were found in a search of other loci with similar properties (*mir-1004–1016*, Supplementary Table S1). The most abundant RNA species from each of the 14 mirtrons, annotated as the mature miRNA, derived from the 3′ arm of its hairpin. Such bias was consistent with the known 5′ nucleotide biases of miRNAs, which frequently begin with a U and rarely with a G (ref. 11). The near-ubiquitous intronic 5′ G, together with other requirements at intron 5′ ends[12], would place unfavourable constraints on miRNAs deriving from the 5′ arm of a mirtron, whereas the species from the 3′ arm would have more freedom. As expected, the species from the 3′ arms, like canonical miRNAs, usually had a 5′ U (12/14 mirtrons).
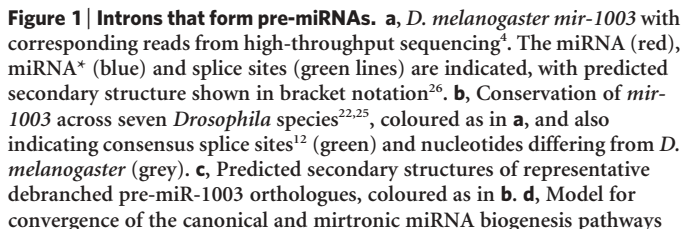
To test whether the small RNAs from mirtrons were functional miRNAs or inactive degradation intermediates, we assessed the gene-silencing capacities of miR-1003 and miR-1006 in *Drosophila* S2 cells. In animals, extensive complementarity leads to cleavage of the target mRNA, but post-transcriptional repression is more commonly mediated by less extensive complementarity, primarily involving pairing to a 5′ region of the miRNA known as the miRNA seed[1]. miR-1003 and miR-1006 repressed reporter genes with perfectly complementary sites, with the repression levels approaching that observed for the *let-7* miRNA and an analogous reporter (Fig. 1e). In addition, both mirtronic miRNAs repressed reporter genes containing *Drosophila* untranslated region (UTR) fragments with seed-based matches typical of metazoan miRNA targets. Conservation of the miR-1003 and miR-1006 seeds (Fig. 1d, Supplementary Table S1) suggested an *in vivo* role for such mirtron-mediated repression; target predictions for conserved mirtronic miRNAs are provided (http://www.targetscan.org).

Having established that mirtrons can direct miRNA-like gene repression, we tested the dependence of mirtron processing on splicing and debranching. A mutant *mir-1003* with a substitution that impaired splicing (3′ Mut) generated little pre- or mature miR-1003 (Fig. 2a, b) and displayed significantly less silencing activity (Fig. 1e). Mutations disrupting the 5′ splice site (5′ Mut) also impaired splicing and miR-1003 accumulation (Fig. 2a, b). Coexpressing a mutant U1 small nuclear RNA (snRNA; U1-3G) that had compensatory changes designed to restore splice site recognition[13] restored splicing of *mir-1003* 5′ Mut (Fig. 2b). Rescuing splicing also restored the levels of pre- and mature miR-1003 (Fig. 2b). These results demonstrated that splicing was required for mirtron maturation and function, which contrasts with the splicing-independent biogenesis of canonical miRNAs found within introns[14].
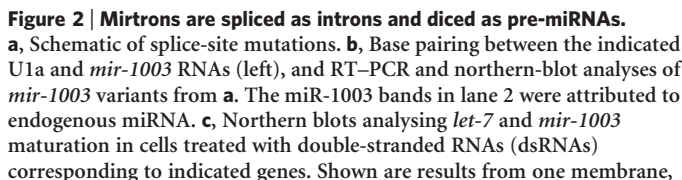
We next used RNA interference (RNAi) knockdown experiments to examine the trans-factor requirements for miR-1003 and miR-1006 biogenesis in *Drosophila* cells. As predicted by our model, in which mirtrons enter the miRNA biogenesis pathway after splicing and debranching, targeting the mRNA of lariat debranching enzyme

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

Figure 1 | Introns that form pre-miRNAs. a, *D. melanogaster mir-1003* with corresponding reads from high-throughput sequencing[4]. The miRNA (red), miRNA* (blue) and splice sites (green lines) are indicated, with predicted secondary structure shown in bracket notation[26]. **b,** Conservation of *mir-1003* across seven *Drosophila* species[22,25], coloured as in **a**, and also indicating consensus splice sites[12] (green) and nucleotides differing from *D. melanogaster* (grey). **c,** Predicted secondary structures of representative debranched pre-miR-1003 orthologues, coloured as in **b**. **d,** Model for convergence of the canonical and mirtronic miRNA biogenesis pathways

(see text). **e,** MicroRNA regulation of luciferase reporters in S2 cells. Plotted is the ratio of repression for wild-type versus mutated sites, normalized to that with the indicated non-cognate miRNA. Bar colour represents the cotransfected miRNA expression plasmid; coloured lines below indicate the cognate miRNA for the specified reporter. Error bars represent the third largest and smallest values from 12 replicates (four independent experiments, each with three transfections; *$P < 0.01$, **$P < 0.0001$, Wilcoxon rank-sum test).

reduced the amount of pre- and mature mirtronic miRNAs without impeding canonical miRNA maturation (Fig. 2c, d). For each mirtron, a probe to the 5′ end of the intron (probe 1) detected both the pre-miRNA hairpin and the accumulating lariat, whereas a probe to the 3′ end of the intron (probe 2) detected the pre-miRNA but failed to detect the lariat, presumably owing to overlap with the branch-point (Supplementary Fig. S1a). Altered relative mobility on gels with different polyacrylamide densities confirmed detection of the

Figure 2 | Mirtrons are spliced as introns and diced as pre-miRNAs. a, Schematic of splice-site mutations. **b,** Base pairing between the indicated U1a and *mir-1003* RNAs (left), and RT–PCR and northern-blot analyses of *mir-1003* variants from **a**. The miR-1003 bands in lane 2 were attributed to endogenous miRNA. **c,** Northern blots analysing *let-7* and *mir-1003* maturation in cells treated with double-stranded RNAs (dsRNAs) corresponding to indicated genes. Shown are results from one membrane,

sequentially stripped and probed for *let-7* RNA, pre-miR-1003/lariat (probe 1), pre-miR-1003/miR-1003 (probe 2), and U6. Previously validated dsRNAs were used[28,29], except for lariat debranching enzyme (CG7942, which we name *ldbr*), for which two unique dsRNAs were used. Knockdowns were confirmed by monitoring mRNA level and protein function (Supplementary Fig. S2). Quantification of band intensities is provided (Supplementary Table S3). *Lariat. **d,** Analysis of *mir-1006* processing, as in **c**.

mirtron lariat (Supplementary Fig. S1b). The debranching knockdown results, together with those of the splice-site mutations and rescue, demonstrated that the intron lariat was an intermediate on the pathway of mirtronic miRNA biogenesis.

Knockdown of other miRNA biogenesis factors further supported our model. As expected if debranched mirtrons enter the later steps of the miRNA pathway rather than the short interfering RNA (siRNA) pathway[3], knockdown of *dicer-1* or its partner, *loquacious*, increased the ratio of pre- to mature mirtronic miRNA, whereas knockdown of *dicer-2* or its partner, *r2d2*, did not (Fig. 2c, d). Knockdown of *drosha* decreased pre- and mature *let-7* RNA accumulation, with little effect on mature miR-1003 or miR-1006 accumulation and a modest effect on mirtronic pre-miRNAs (Fig. 2c, d). The more modest effect on mirtronic pre- and mature miRNAs supported the idea that mirtronic pre-miRNAs are not Drosha cleavage products. The decrease of mirtronic pre-miRNA that was observed would be explained if Drosha bound mirtronic pre-miRNAs, stabilized them from degradation, and perhaps facilitated their loading into the nuclear export machinery. The decrease could also reflect increased Dicer-1 accessibility in the *drosha* knockdown due to reduced substrate competition from endogenous pre-miRNAs. In this case, simultaneous knockdown of *dicer-1* and *drosha* would lead to a more substantial accumulation of pre-miRNAs derived from mirtrons than from canonical miRNAs, as was observed for pre-miR-1003 and pre-miR-1006 compared to *let-7* pre-miRNA (Fig. 2c, d).

The distribution of intron lengths, which varies widely in different organims[12,15], would influence the probability of new mirtrons arising during evolution. The introns of *Drosophila* share a similar length distribution with the annotated pre-miRNAs, producing a context particularly well suited to the emergence to mirtrons (Fig. 3a, c). *C. elegans* also has a substantial number of pre-miRNA-sized introns. Indeed, examination of prior miRNA annotations revealed that *mir-62*, which produces a highly conserved nematode miRNA that was among the very first to be cloned in animals[11,16], had mirtron-like properties (Fig. 3b). Like the mirtrons of *D. melanogaster*, the base pairing capacity of the sequence surrounding pre-miR-62 ended at the border of the host intron, and the most abundant miRNA 3′ terminus corresponded to the 3′ splice site (with the single read whose 3′ terminus extended into the 3′ exon attributable

to untemplated nucleotide addition to the miRNA 3′ end[5]). A directed search of *C. elegans* small RNA sequences[5] revealed three more mirtrons, annotated here as *mir-1018–1020* (Supplementary Table S2).

Even if only a very small portion of debranched introns can form secondary structures resembling those of pre-miRNAs, the abundance of pre-miRNA-sized introns in flies and nematodes would allow a large absolute number of candidate mirtrons to emerge over evolutionary timescales. Whether they persist as functional mirtrons depends on the selective advantage conferred to the host organism as a consequence of their gene-repression activities. This model for mirtron emergence predicts that, at any historical point, some introns will be processed as mirtrons that provide no advantage to the organism but have yet to be eliminated by natural selection or neutral drift. Accordingly, some but not all processed *D. melanogaster* mirtrons were significantly more conserved in *Drosophila pseudoobscura* than were most small introns, and the same trend was observed for *C. elegans* mirtrons in *Caenorhabditis briggsae* (Fig. 3d). The three most conserved *D. melanogaster* mirtrons (*mir-1003/1006/1010*) gave rise to more reads than 27%, 16% and 4% of the non-mirtronic miRNAs conserved to *D. pseudoobscura*, respectively[4], while the most conserved *C. elegans* mirtron (*mir-62*) gave rise to more reads than 52% of the non-mirtronic miRNAs conserved to *C. briggsae*[5].

Compared to flies and nematodes, mammals have few pre-miRNA-sized introns[12,15] (Fig. 3a), perhaps explaining why we found no mirtrons among the annotated mammalian miRNAs[17]. Nonetheless, high-throughput sequencing of mammalian small RNAs might yet reveal mirtrons. In plants, miRNA processing could similarly bypass one of the RNase III cleavages, although plant mirtrons have not yet been identified[1,17]. Moreover, lineages with long introns might have other types of intronic miRNAs that bypass Drosha-mediated cleavage. This possibility was raised by *mir-1017*, whose putative pre-miRNA 5′ end, but not 3′end, matched the 5′ splice site of its host intron (Supplementary Table S1). In contrast to true mirtrons, miRNAs of this type would depend on a nuclease to cleave their extensive 3′overhangs, as observed for the U14 snRNA derived from an intron of *hsc70* (ref. 18). This mechanism, together with that of mirtron processing, would enable miRNAs to emerge in any organism with both splicing and post-transcriptional RNA silencing,
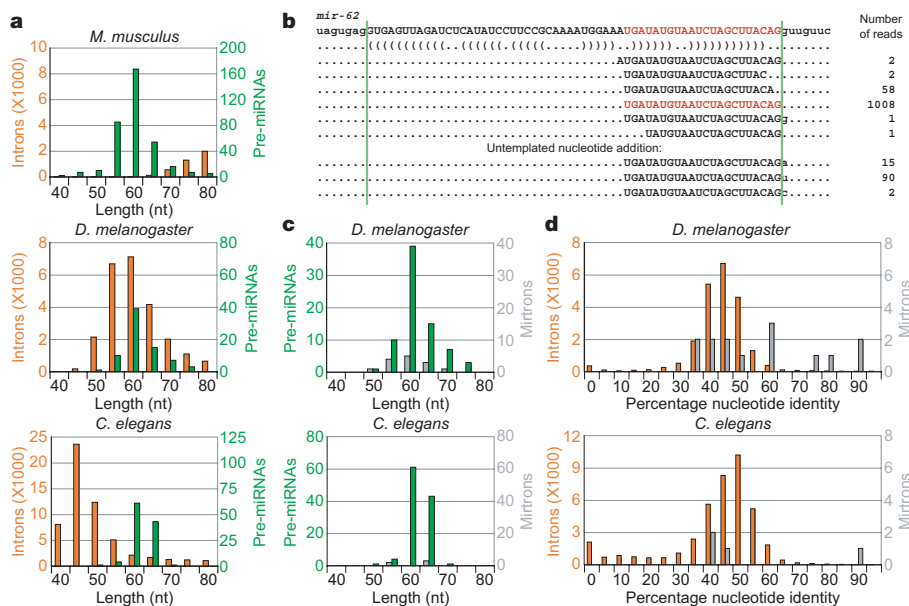

Figure 3 | Emergence and conservation of mirtrons in species with appropriately sized introns. a, Distributions of intron (orange) and pre-miRNA (green) lengths from the indicated species. Introns and pre-miRNAs were binned by length. **b**, Intron and associated reads of *C. elegans* mir-62 (ref. 5), coloured as in Fig. 1a. Reads with untemplated nucleotides added at

their 3′ terminus are shown below. **c**, Distributions of pre-miRNA (green) and mirtron (grey) lengths from *D. melanogaster* and *C. elegans*. **d**, Conservation of all 40–90-nt introns (orange) versus mirtrons (grey) from *D. melanogaster* (percentage identity shared with *D. pseudoobscura*) and *C. elegans* (percentage identity shared with *C. briggsae*).

even those lacking the specialized RNase III enzyme Drosha or its plant counterpart, DICER-LIKE1 (ref. 1). In this scenario, miRNAs might have emerged in ancient eukaryotes before the advent of modern miRNA biogenesis pathways.

METHODS SUMMARY

Computational methods. *D. melanogaster* small RNAs were from 2,075,098 high-throughput pyrosequencing reads[4] and are available at the GEO. *C. elegans* small RNA sequences were from ref. 5. Introns were as annotated in FlyBase (v4.2)[19], WormBase (release WS120)[20] and human RefSeq annotations[21] available through UCSC (hg17)[22]. Percentage conservation of *D. melanogaster*[23] and *D. pseudoobscura*[24] introns was calculated as the number of identity matches between the two orthologous introns in the multiZ alignment[22,25] divided by the length of the longer intron. *C. elegans* intron conservation was similarly determined using multiZ alignments[22] of the *C. elegans* and *C. briggsae* (WormBase cb25.agp8) genomes[20,22]. Pre-miRNA lengths were the sum of the miRNA length, the miRNA* length, and the length of intervening sequence, calculated after using RNAfold[26] to predict the structure of annotated miRNA hairpins (miRBase v9.1)[17] and inferring the miRNA* by assuming 2-nt 3′ overhangs when paired with the annotated miRNA.

Analysis of function and biogenesis. Mirtron minigenes containing flanking exons were amplified from genomic DNA and cloned into expression vectors, pMT-puro or p2032 (ref. 27). Similar plasmids were constructed for a 780-base-pair (780-bp) fragment centred on the *let-7* hairpin. Luciferase reporters were constructed with 3′ UTRs (Supplementary Table S3) amplified from genomic DNA. U1 plasmids were constructed as described[13]. Mutations to seed sites (reporters) or splice sites (minigenes) were introduced by Quikchange site-directed mutagenesis (Stratagene). After RNAi knockdown[28,29], miRNA expression was induced with 500 μM CuSO₄, then 12 h post-induction RNA was extracted with TRI reagent and analysed on northern blots[5]. *Renilla* (reporter) and firefly (control) luciferase plasmids were cotransfected with miRNA-expressing plasmid into S2 cells. Fold repression was calculated by dividing normalized luciferase activity for mutant reporters by that of wild-type reporters in the presence of cognate miRNA. Transfection with non-cognate miRNA served as a specificity control.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
2. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
3. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005).
4. Ruby, J. G. *et al.* Biogenesis, expression, and target predictions for an expanded set of microRNA genes in *Drosophila*. *Genome Res.* (in the press)..
5. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
6. Lim, L. P. *et al.* The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* **17**, 991–1008 (2003).
7. Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).
8. Ruskin, B. & Green, M. R. An RNA processing activity that debranches RNA lariats. *Science* **229**, 135–140 (1985).
9. Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. & Sharp, P. A. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* **225**, 898–903 (1984).
10. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
11. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
12. Lim, L. P. & Burge, C. B. A computational analysis of sequence features involved in recognition of short introns. *Proc. Natl Acad. Sci. USA* **98**, 11193–11198 (2001).
13. Lo, P. C., Roy, D. & Mount, S. M. Suppressor U1 snRNAs in *Drosophila*. *Genetics* **138**, 365–378 (1994).
14. Kim, Y. K. & Kim, V. N. Processing of intronic microRNAs. *EMBO J.* **26**, 775–783 (2007).
15. Yandell, M. *et al.* Large-scale trends in the evolution of gene structures within 11 animal genomes. *PLoS Comput. Biol.* **2**, e15 (2006).
16. Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
17. Griffiths-Jones, S. The microRNA registry. *Nucleic Acids Res.* **32**, D109–D111 (2004).
18. Leverette, R. D., Andrews, M. T. & Maxwell, E. S. Mouse U14 snRNA is a processed intron of the cognate hsc70 heat shock pre-messenger RNA. *Cell* **71**, 1215–1221 (1992).
19. Grumbling, G. & Strelets, V. FlyBase: anatomical data, images and queries. *Nucleic Acids Res.* **34**, D484–D488 (2006).
20. Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J. & Spieth, J. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.* **29**, 82–86 (2001).
21. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
22. Karolchik, D. *et al.* The UCSC genome browser database. *Nucleic Acids Res.* **31**, 51–54 (2003).
23. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
24. Richards, S. *et al.* Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* **15**, 1–18 (2005).
25. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
26. Hofacker, I. L. *et al.* Fast folding and comparison of RNA secondary structures. *Mh. Chem.* **125**, 167–188 (1994).
27. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
28. Dorner, S. *et al.* A genomewide screen for components of the RNAi pathway in *Drosophila* cultured cells. *Proc. Natl Acad. Sci. USA* **103**, 11880–11885 (2006).
29. Forstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
30. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).

METHODS

Computational methods. *D. melanogaster* small RNAs were from 2,075,098 high-throughput pyrosequencing reads[4] and are available at the GEO. *C. elegans* small RNA sequences were from ref. 5. Introns were defined according to FlyBase v4.2 *D. melanogaster* gene annotations[19]. *C. elegans* introns were defined using annotations and genomic sequence from WormBase (release WS120)[20]. *Mus musculus* introns were defined using NCBI RefSeq annotations[21] applied to the March 2005 release of the mouse genome available through UCSC (mm6)[22]. RNA secondary structures were predicted using RNAfold[26]. *D. melanogaster* intron conservation was assessed based on a nine-species multiZ alignment[25] of *D. melanogaster*, *Drosophila simulans*, *Drosophila yakuba*, *Drosophila ananassae*, *D. pseudoobscura*, *Drosophila virilis*, *Drosophila mojavensis*, *Anopheles gambiae* and *Apis mellifera* genomes, generated at UCSC[22]. Percentage nucleotide identity between *D. melanogaster* and *D. pseudoobscura* introns was calculated as the number of identity matches between the two orthologous introns in the multiZ alignment divided by the length of the longer intron. Introns not aligned between those two species were not tallied. *C. elegans* intron conservation was similarly determined using multiZ alignment of the *C. elegans* and *C. briggsae* (WormBase cb25.agp8)[20] genomes generated at UCSC[22]. Pre-miRNA lengths were calculated using miRBase v9.1 hairpin annotations[17]. Secondary structures were generated using RNAfold[26], and the miRNA* position was inferred on the basis of the annotated miRNA, assuming 2-nt 3′ overhangs. Pre-miRNA lengths were the sum of the miRNA length, the miRNA* length, and the length of intervening sequence.

Plasmids. Minigenes containing *mir-1003* and *mir-1006* and flanking exons were PCR amplified from genomic DNA. Minigenes for *mir-1006* and *mir-1003* were cloned into pMT-puro with the indicated sites to make expression plasmids pCJ19 and pCJ20, respectively. *let-7* was amplified from genomic DNA with primers 474 bp upstream and 310 bp downstream of the *let-7* hairpin and cloned into pMT-puro to make pCJ24. Similar minigenes replaced EGFP in p2032 (ref. 27) to give pCJ31 (*mir-1006*), pCJ30 (*mir-1003*) or pCJ32 (*let-7*). U1a snRNA and U1a-3G snRNA expression constructs were constructed essentially as described[13]. Sequences of inserts in pCJ19 (pMT-puro-*mir-1006*), pCJ20 (pMT-puro-*mir-1003*), pCJ24 (pMT-puro-*let-7*), pCJ30 (p2032-*mir-1003*), pCJ31 (p2032-*mir-1006*), and pCJ32 (p2032-*let-7*) are provided (Supplementary Table S4). Quikchange site-directed mutagenesis (Stratagene) was used to make 3′ splice site mutations with the indicated primers: *mir-1003* 3′ mut (CCTCTCACATTTACATATTCACGACGCCGTGAGCTGC and GCAGCTCACGGCGTCGTGAATATGTAAATGTGAGAGG), and *mir-1006* 3′ mut (GGTACAATTTAAATTCGATTTCTTATTCATGCGTGCAATACCAGTTGATC and GATCAACTGGTATTGCACGCATGAATAAGAAATCGAATTTAAATTGTACC). Similarly, *mir-1003* 5′ mut was made with the following mutagenic primers: (GCTGCGCAGAACGTGGGCATCTGGATGTGGTTGGC and GCCAACCACATCCAGATGCCCACGTTCTGCGCAGC; CCTCTCACATTTACATGTTCACAGGCGCCGTGAG and CTCACGGCGCCTGTGAACATGTAAATGTGAGAGG).

Luciferase-reporter inserts were made by annealing oligonucleotides with their reverse complements, leaving overhangs for the indicated restriction sites (lower case): let-7-ps (gagctcACTATACAACCTACTACCTCAaactagt), let-7-psm (gagctcACTATACAACCTACAAGCACAaactagt), miR-1003-ps (gagctcCTGTGAATATGTAAATGTGAGAaactagt), miR-1003-psm (gagctcCTGTGAATATGTAAAAGAGTGAactagt), miR-1006-ps (gagctcCTATGAATAAGAAATCGAATTTAactagt), and miR-1006-psm (gagctcCTATGAATAAGAAATCCATTATAactagt). Annealed oligos were ligated into *Sac*I/*Spe*I-cleaved pIS2 (ref. 30). These plasmids were linearized with *Hind*III, polished with Klenow enzyme to create blunt ends, and digested with *Not*I to excise the *Renilla* luciferase gene with the modified UTR from the remainder of pIS2. The gel-purified *Renilla* gene fragment was then ligated into pMT-puro between *Eco*RV and *Not*I sites for copper-induced expression in S2 cells.

Cell culture and RNAi. S2-SFM cells were adapted from S2 cells to grow in *Drosophila* serum free media (SFM) by passaging into increasing amounts of SFM (0%, 25%, 50%, 75%, 90%, 100%), then grown in SFM supplemented with 2 mM L-glutamine at 25 °C in a humidified incubator. 5 µg of pCJ19 or pCJ20 were transfected into a 60 mm plate containing 2.5×10^6 S2 cells with FuGENE HD. Cells were grown for 3 days, split 1:10, and selected for 3 weeks in 10 µg ml^{-1} puromycin before experimentation, then maintained in 5 µg ml^{-1} puromycin.

Templates for dsRNA were amplified by PCR and extended to have convergent T7 promoters. 400 µl PCR reactions were phenol/chloroform extracted, ethanol precipitated, and used as template for 400 µl T7 transcriptions. Transcription reactions were treated with 20 U of DNase I for 15 min. The transcription products were then extracted in phenol:chloroform (5:1 pH 5.3) and ethanol precipitated. RNA was resuspended, desalted over Sephadex G-300, then heated to 75 °C for 10 min and slow cooled to room temperature. Yield and quality were assessed by agarose gel and UV absorbance. The sense sequence of each dsRNA is listed (Supplementary Table S4).

S2 cells were soaked in 10 µg ml^{-1} dsRNA in SFM. 500,000 cells were plated per well of a 24-well plate and soaked for 2 days, split 1:4, soaked another 2 days, expanded into 6-well plates, then soaked for three days. MicroRNA expression was induced by addition of 500 µM CuSO$_4$ to the growth media, and RNA harvested 12 h later with TRI reagent.

Northern blots were performed as described[5], using the following oligonucleotides (purchased from IDT) as probes for the indicated RNA species ('+' precedes LNA bases): ACTATACAACCTACTACCTCA (*let-7*), C+TGT+GAA+TAT+GTA+AAT+GTG+AGA (*mir-1003* probe 1), CCAACCACATCCAGATACCCACC (*mir-1003* probe 2), C+TAT+GAA+TAA+GAA+ATC+GAA+TTT+A (*mir-1006* probe 1), TTTACGCATTTCAATTTCAAACTCAC (*mir-1006* probe 2), TTGCGTGTCATCCTTGCGCAGG (U6).

RT–PCR. 500 ng mirtron plasmids were cotransfected with 500 ng either U1 or GFP carrier plasmid using 3 µl FuGENE HD per well of a 12 well plate. 24 h post-transfection, mirtron expression was induced for 36 h in the presence of 500 µM CuSO$_4$. Total RNA was extracted with TRI reagent, and 4 µg were treated with DNase using the DNA-free kit (Ambion). 500 ng DNA-free RNA were reverse-transcribed with oligo-dT(16) and Superscript III (Invitrogen) per manufacturers instructions. 1 µl cDNA was used as a template for PCR using exonic primers (ATAAAGCCGATAAGCGTGCG and CGTCCTTGTGCGTCTCC-TCC) flanking *mir-1003*. After 24 cycles of PCR, 10 µl of the reaction was resolved on an ethidium-stained 1.5% agarose gel and visualized by UV illumination.

Quantitative RT–PCR was performed on an ABI 7000 Real-Time PCR system with ABI Power SYBR Green reagents. First-strand synthesis was performed as above. The following primer pairs were used to amplify the specified mRNA: *actin 5c* (CCCATCTACGAGGGTTATGC, TTGATGTCACGGACGATTTC); *drosha* (TCACCATCCACGAGCTAGAC, ACGAAACGCGGAAAGAAGTG); *dicer-1* (GCCATTGAAGCATGACATTG, AAATCCCTCCTTGCCGATAG); *loquacious* (CGATTACCGAGTGGATACGG, CAAAGGAATCGGTGGAAAAG); *dicer-2* (GGCCACGAAACTTAAAGAGC, TGTGGAAAGGACACCATGAC); *r2d2* (GACGGAGGGTACGTCTGTAAA, AGCAGTTGGATTTTACGCAAG); *ldbr* (TTATCCCTGCCAGCACCTAC, CCTCTACATGAGGCGTTTCC).

Threshold cycle (Ct) and baseline were detected by ABI 7000 SDS software. *actin 5C* was used to calculate the ΔCt, and ΔΔCt was calculated by subtracting the ΔCt from that of the GFP dsRNA treated samples; the relative abundance was calculated as $1/(2^{\Delta\Delta Ct})$. Geometric mean ± standard deviation are shown for three replicate wells.

Luciferase assays. S2-SFM cells were plated 300,000 cells ml^{-1} in 96 well plates. After 24 h, cells were cotransfected with 96 ng microRNA-expressing plasmid, 4 ng perfect-site reporter and 2 ng firefly reporter per well using FuGENE HD (3 µl lipid per µg DNA). Expression of *Renilla* luciferase was induced 24 h post-transfection with 500 µM CuSO$_4$. Luciferase assays were performed 24 h post-induction with the Dual-Glo Luciferase system (Promega) on a Tecan Safire2 plate reader. The ratio of *Renilla*:firefly luciferase activity was measured for each well. To calculate fold repression, the ratio of *Renilla*:firefly for reporters with mutant sites was divided by the ratio of *Renilla*:firefly for reporters with wild-type sites. These values were also obtained in the presence of a plasmid expressing a non-cognate miRNA, and fold repression for the cognate miRNA was normalized to that of the non-cognate.

Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity

Stanley S. Ng[1], Kathryn L. Kavanagh[1], Michael A. McDonough[2], Danica Butler[2], Ewa S. Pilka[1], Benoit M. R. Lienard[2], James E. Bray[1], Pavel Savitsky[1], Opher Gileadi[1], Frank von Delft[1], Nathan R. Rose[2], John Offer[3], Johanna C. Scheinost[3], Tomasz Borowski[4], Michael Sundstrom[1], Christopher J. Schofield[2] & Udo Oppermann[1]

Post-translational histone modification has a fundamental role in chromatin biology and is proposed to constitute a 'histone code' in epigenetic regulation[1,2]. Differential methylation of histone H3 and H4 lysyl residues regulates processes including heterochromatin formation, X-chromosome inactivation, genome imprinting, DNA repair and transcriptional regulation[3]. The discovery of lysyl demethylases using flavin (amine oxidases)[4] or Fe(II) and 2-oxoglutarate as cofactors (2OG oxygenases)[5–7] has changed the view of methylation as a stable epigenetic marker. However, little is known about how the demethylases are selective for particular lysyl-containing sequences in specific methylation states, a key to understanding their functions. Here we reveal how human JMJD2A (jumonji domain containing 2A), which is selective towards tri- and dimethylated histone H3 lysyl residues 9 and 36 (H3K9me3/me2 and H3K36me3/me2), discriminates between methylation states and achieves sequence selectivity for H3K9. We report structures of JMJD2A–Ni(II)–Zn(II) inhibitor complexes bound to tri-, di- and monomethyl forms of H3K9 and the trimethyl form of H3K36. The structures reveal a lysyl-binding pocket in which substrates are bound in distinct bent conformations involving the Zn-binding site. We propose a mechanism for achieving methylation state selectivity involving the orientation of the substrate methyl groups towards a ferryl intermediate. The results suggest distinct recognition mechanisms in different demethylase subfamilies and provide a starting point to develop chemical tools for drug discovery and to study and dissect the complexity of reversible histone methylation and its role in chromatin biology.

The 2OG-dependent histone demethylases belong to a ubiquitous family of oxygenases that catalyse a diverse set of reactions including hydroxylations, C–C bond cleavages and halogenations[8] (Supplementary Fig. 1). Full-length JMJD2A protein (1,064 residues) comprises a catalytic core composed of JmjN and JmjC domains, a ~350-amino-acid linker region, two plant homeodomains (PHD), and a double Tudor domain. Crystal structures of the catalytic domain of the histone demethylase JMJD2A in complex with 2OG revealed the active-site Fe(II) coordinated by the His 188, Glu 190 and His 276 side chains, and in a bidentate manner the 2OG[9]. These structures defined the anticipated[10] double-stranded β-helical fold and a novel Zn-binding site, but did not reveal how the methylated lysyl substrates are bound to the active site or how selectivity is achieved. To address these questions we solved the structures of JMJD2A in complex with a set of histone H3K9 and H3K36 substrates and products of different lysyl methylation states.

Initially we investigated the selectivity of JMJD2A (Supplementary Table 1 and Supplementary Figs 2–3). In accord with previous studies[6,7], we found that JMJD2A catalyses the demethylation of trimethylated H3K9, and also of dimethylated H3K9 and trimethylated H3K36 peptides, albeit at lower rates. Methylated H3K4 and H3K27 peptides were not converted; Nε-acetylation of H3K14 did not affect activity. These analyses enabled identification of suitable substrates for co-crystallization: an H3K9 octapeptide (Ala 7–Lys 14) and an H3K36 undecapeptide (Ser 31–Tyr 41). Crystals of JMJD2A were obtained under aerobic conditions with three different H3 substrates and a monomethylated product, comprising peptides that were mono-, di- and trimethylated at Lys 9 and trimethylated at Lys 36. Two essential components were added to form a catalytically inert complex: Ni(II) replacing the endogenous Fe(II), and a 2OG competitive inhibitor, N-oxalylglycine (NOG), replacing 2OG.

The refined structures with resolutions of 1.95–2.3 Å (Supplementary Table 2) reveal that in the peptide complexes the JMJD2A backbone adopts a similar conformation as observed in the Fe(II)–2OG complex, indicating that significant conformational changes, as proposed[9], are not required for substrate binding. There are two protein monomers in the asymmetric unit, and well defined electron density for the substrate peptides—except for the His 39 and Arg 40 side chains, and for Ser 31—allowed model building over their entire length. The side chains of the Zn(II)-binding residues His 240, Cys 306, Cys 308 and Cys 234 do not directly interact with the substrate peptide; however, the His 240 backbone carbonyl is positioned to form a hydrogen bond with the H3 Gly 13 backbone NH (~3.4 Å). Residues Lys 241 and Arg 309, adjacent to the Zn-coordinating residues His 240 and Cys 308 respectively, are involved in substrate binding, indicating a role for the Zn-binding site in peptide recognition (Supplementary Figs 5 and 6).

Each substrate peptide bends around a flexible loop surrounding Asp 311 located between helices α9 and α10; this conformation projects the H3K9 or H3K36 methylated side chain towards the active-site metal (Fig. 1; for stereoviews and electron density see Supplementary Fig. 5). Molecular dynamics simulations of H3K9 substrates for both Ni(II) and Fe(II) metal centres predict that the overall structure of the complexes, in particular the cationic side chain, will be the same in solution as in the crystalline forms (see Supplementary Fig. 7).

Despite their different activities (Supplementary Table 1), the positions of the H3K9me3 and H3K36me3 lysyl side chains were superimposable; however, binding for the two substrates is achieved through distinct conformations. The H3K9 substrate binds

[1]Structural Genomics Consortium, Botnar Research Center, University of Oxford, Oxford OX3 7LD, UK. [2]The Department of Chemistry and Oxford Centre of Integrated Systems Biology, University of Oxford, Mansfield Road, Oxford OX1 3TA, UK. [3]The Scripps-Oxford Laboratory, The Department of Biochemistry, University of Oxford, Mansfield Road, Oxford OX1 3QU, UK. [4]Institute of Catalysis and Surface Chemistry, PAS University, Niezapominajek 830-239 Cracow, Poland.
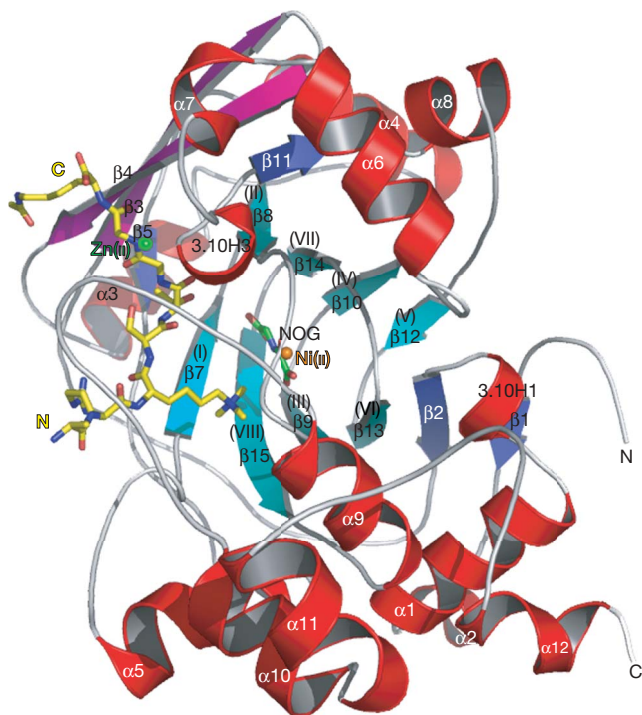
Figure 1 | Overview of the H3K9 peptide bound to the catalytic domain of JMJD2A. Ribbon model of the JMJD2A–Ni(II)–Zn(II)–NOG–H3K9me3 complex with secondary structure in the double-stranded β-helical core elements labelled I–VIII. The double-stranded β-helical core (cyan) of JMJD2A is extended by β3 and β4 (magenta) at the H3K9me3 substrate-binding cavity. JMJD2A is coloured with helices in red and other β-strands in blue.

approximately in a broad 'W'-shaped conformation in which residues downstream of Lys 9 interact with the substrate pocket (Fig. 2). The bound H3K36 adopts a tighter bend than in the H3K9 substrate, leading to a 'U'-shaped binding conformation (Fig. 2). With the H3K9 substrate, the H3 Ser 10 side chain and main-chain H3 Gly 12 amide are positioned to form an intra-substrate hydrogen bond that will stabilize the bent conformation. Evidence that this hydrogen bond is important for productive binding comes from the observation that a Ser10Ala mutant of a H3K9me3 peptide showed strongly reduced activity (Fig. 3). Furthermore, phosphorylation at Ser 10 ablates activity (Fig. 3), consistent with the proposal that phosphorylation and methylation states represent a 'switch' mechanism between gene silencing and activation[11].

A comparison between H3K9 and H3K36 peptides reveals that the amide link between Lys 36 and Lys 37 is rotated 180° relative to that of Lys 9–Ser 10. The Lys 37–Pro 38 peptide bond is in the *trans* configuration. In contrast to a previous proposal that a *cis*-proline would provide the necessary bent conformation for peptide binding[12], inspection of the structure suggests that binding of a peptide containing a *cis*-proline would be disfavoured, as this would place the Lys 37–Pro 38 amide carbonyl adjacent to Asp 135. Competition experiments between wild-type and Pro38Ala H3K36 revealed that the mutant was converted at a similar rate to the wild type (Supplementary Fig. 4), demonstrating that a prolyl residue at this position is not essential and arguing against a requisite *cis*-amide bond between residues 37 and 38. Inter-conversion (*trans* to *cis*) of the Lys 37–Pro 38 peptide bond configuration by the Fpr4 peptidyl proline isomerase antagonizes Set2-mediated H3K36 methylation in yeast[12], indicating a crosstalk between non-covalent modifications (prolyl isomerization) and covalent modifications (methylation) in regulation. The lack of a definitive proline specificity

Figure 2 | JMJD2A substrate peptide binding. a, The H3K9me3 peptide in relation to the active site metal Ni(II) (in place of Fe(II)) and co-substrate analogue NOG. **b**, The H3K36me3 peptide conformation. **c**, **d**, Ligplot representations of H3K9me3 (**c**; centre, yellow) and H3K36me3 (**d**; centre,

blue) binding. Because electron density was not observed for part of Ser 31 and the side-chains of His 39 and Arg 40, the atoms were not included in the final model. Distances (dashed lines) are given as ångstroms.

determinant in H3K36 recognition suggests a certain degree of plasticity in substrate recognition for JMJD2A and related subfamily members.

We conclude that binding of methylated lysyl peptides to the active site is achieved through few polar side-chain interactions, unusual for a sequence-specific enzyme, and requires a bent peptide conformation. For the preferred H3K9 substrate the required conformation is apparently enabled by a Gly–Gly sequence, with an intra-substrate hydrogen bond to Ser 10 resulting in a wider arc than in H3K36. Replacement of the Gly–Gly motif in H3K9 with two alanine residues results in complete loss of activity (Fig. 3 and Supplementary Table 1), whereas a H3K27 peptide with a AARK(me3)SAGAT sequence (instead of AARK(me3)SAPAT as in wild-type H3K27) gains activity. Transposition of the Ala–Pro motif of H3K27 to AARK(me3)SPAAT, to mimic the proline position in H3K36, resulted in no detectable activity. These results highlight the importance of the correct positioning of Pro and Gly residues in relation to the methylated Lys for activity, but also that surrounding residues are important in determining selectivity. Together, the results rationalize the lack of demethylase activity towards H3K27, which has an identical −2 (relative to the methylated lysyl) upstream sequence to H3K9 but has a prolyl residue at position +3 (Supplementary Fig. 8); this may force the H3K27 peptide to clash with the active-site wall, thus disallowing a productive binding conformation.

Other sequences around methylated lysyl residues (H3K4, H3K79 and H4K20) have dissimilar upstream sequences, but lack putative bend-enabling Gly-rich sequences, possibly rationalizing the lack of demethylase activity towards them[6,7] (Supplementary Fig. 8). A sequence alignment of human 2OG demethylases suggests that distinct mechanisms of histone substrate recognition occur for different demethylase subfamilies: the JMJD2A residues involved in peptide binding are not conserved across the different demethylase subfamilies, whereas residue conservation in Fe(II) and co-substrate binding is observed (Supplementary Fig. 9).

Electron density was well defined for all three methyl groups in the H3K9me3 and H3K36me3 structures (a, b and c in Fig. 4a), where the N–C bond of 'a' projects towards Asn 290, 'b' towards Ni(II), which substitutes for Fe(II), and 'c' towards Tyr 177. Although movement may occur in catalysis, these observations support a mechanism in which hydroxylation of the specific methyl group gives a hemiaminal, mediated by a ferryl intermediate, as observed for other 2OG oxygenases[13,14] (Fig. 4b). In both structures with trimethylated peptides, the side chains of Tyr 177, Glu 190, Ser 288 and Asn 290, and the carbonyl of Gly 170, point towards the positively charged methylated nitrogen. One edge of the Tyr 175 phenyl ring lies along the substrate lysyl methylenes and is probably important in productive substrate positioning. One of the Glu 190 carboxylate oxygens coordinates Ni(II) whereas the other is directed towards the trimethyl group. Participation of the Fe(II)-binding acidic residue in substrate coordination is similar to that seen for the Asn-hydroxylase FIH[15], the closest homologue to JMJD2A in the Protein Data Bank.

We also solved H3K9 peptide structures in the di- and mono-methyl states; in each, the overall peptide binding is similar to the H3K9me3 form (Supplementary Fig. 10). Notably, the lone methyl group of the monomethyl lysine occupies position a, that is, it is directed away from the metal (methyl C to Ni distance: >5 Å). In the monomethyl product structure, two additional water molecules are observed close to the positions occupied by the b and c methyl groups in the trimethyl structure. The water molecule in the c site is within hydrogen-bonding distance of the Gly 170 and Ser 288 main-chain carbonyls, whereas the water molecule in the b site could interact with the hydroxyls of Tyr 177 and Ser 288 and the water involved in metal coordination. These water molecules are positioned to direct the single methyl group away from the hydroxylation site adjacent to the metal ion, consistent with the lack of activity with a monomethylated peptide.

In the case of the H3K9me2 structure, the dimethyl lysine side-chain is observed primarily in a catalytically non-productive conformation with methyl groups occupying positions a and c (Supplementary Fig. 10). The electron density for the bound peptide indicates two possible conformations; the lower occupancy one adopts a catalytically productive orientation with methyl groups occupying sites a and b. In each conformation a single water molecule—similarly positioned to one of the water molecule positions in the monomethyl structure—occupies the position of the absent methyl group. The lower occupancy of the productive conformation may help to explain the significantly lower activity with the dimethyl substrate.

Overall, our study reveals the basis for H3K9 substrate recognition and methyl selectivity of a 2OG-dependent histone lysyl demethylase. The biological role of histone lysyl methylation at specific residues is a function both of the residue environment and its methylation state. Sequence analyses suggest that there are about 20–30 putative histone demethylating enzymes within an individual mammalian genome. So far, demethylating activities towards histone residues H3K4, H3K9 and H3K36 have been identified[5–7,16,17], however it is likely that other 2OG demethylases will catalyse the demethylation of other methylated histone lysyl and, possibly, arginyl residues.
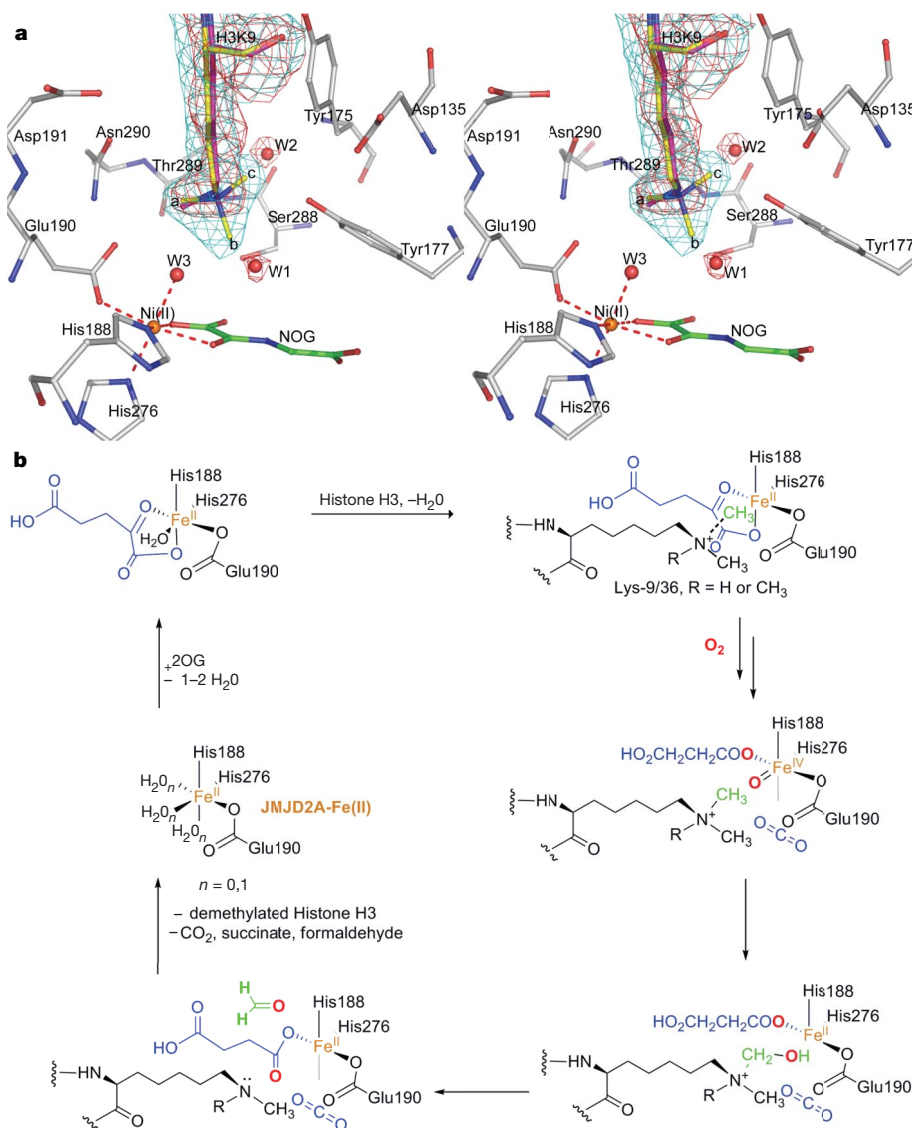

Figure 3 | Mass spectra (MALDI–TOF) of JMJD2A demethylation reactions with native and modified histone peptides. **a–d**, H3K9; **e**, **f**, H3K27; **g**, **h**, H3K36. **a**, H3K9me3; **b**, H3K9me3 Gly12Ala–Gly13Ala; **c**, H3K9me3 Ser 10 phosphorylated; **d**, H3K9me3 Ser10Ala; **e**, H3K27me3; **f**, H3K27me3 Pro30Gly (glycine introduced as in H3K9 sequence); **g**, H3K36me3; **h**, H3K36me3 Pro38Ala.

Figure 4 | Active site and mechanism of JMJD2A. a, Stereoview omit $2F_o - F_c$ electron density maps contoured at 1σ for H3K9me3 (blue) and H3K9me1 (red). The positions (a, b, c) for the three methyl groups of the H3K9me3 substrate are indicated with b projecting towards the metal-bound water. In the H3K9me1 structure the methyl group occupies position a with two water molecules (W1 and W2) close to the b and c positions. **b,** Outline of the catalytic cycle for JMJD2A showing the proposed ferryl and hemiaminal intermediates. Oxygen may bind at the position *trans* to His 276 or His 188.

METHODS SUMMARY

Recombinant protein expression and demethylation assays. A truncated construct encompassing the JmjN and JmjC domains of human JMJD2A was expressed as an amino-terminally His-tagged protein in *Escherichia coli* and purified by metal affinity, size exclusion and anion exchange chromatography. Enzymatic activity was analysed by incubating purified recombinant protein in the presence of Fe(II), 2OG and peptide. Demethylated peptide product formation was assessed by matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry.

Crystallography. NOG was added to the protein solution to a final concentration of 10 mM before setting up drops. For crystallization with peptides (denoted in Supplementary Table 1), 5 mM peptide was included. Crystals were grown by vapour diffusion at 4 °C with 0.1 M citrate pH 5.5, 20% PEG 3350 and 4 mM NiCl₂. Data sets were collected at the European Synchrotron Radiation Facility (ESRF), and the Swiss Light Source (SLS), and on a Rigaku FR-E generator. Crystals were isomorphous with the previously determined JMJD2A structure (Protein Data Bank 2GP3), and initial phases for the Ni-NOG data set were calculated using this structure minus ligands and solvent. The refined Ni-NOG structure was used as a starting model for building the peptide complexes. Statistics are given in Supplementary Table 2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Margueron, R., Trojer, P. & Reinberg, D. The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.* **15**, 163–176 (2005).
2. Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360 (2001).
3. Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nature Rev. Mol. Cell Biol.* **6**, 838–849 (2005).
4. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
5. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–816 (2006).
6. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nature Rev. Genet.* **7**, 715–727 (2006).
7. Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467–481 (2006).
8. Flashman, E. & Schofield, C. J. The most versatile of all reactive intermediates? *Nature Chem. Biol.* **3**, 86–87 (2007).

9. Chen, Z. et al. Structural insights into histone demethylation by JMJD2 family members. *Cell* **125**, 691–702 (2006).
10. Clifton, I. J. et al. Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins. *J. Inorg. Biochem.* **100**, 644–669 (2006).
11. Lo, W. S. et al. Snf1–a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. *Science* **293**, 1142–1146 (2001).
12. Nelson, C. J., Santos-Rosa, H. & Kouzarides, T. Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell* **126**, 905–916 (2006).
13. Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. & Krebs, C. Kinetic dissection of the catalytic mechanism of taurine:α-ketoglutarate dioxygenase (TauD) from *Escherichia coli. Biochemistry* **42**, 7497–7508 (2003).
14. Grzyska, P. K. et al. Steady-state and transient kinetic analyses of taurine/α-ketoglutarate dioxygenase: effects of oxygen concentration, alternative sulfonates, and active-site variants on the FeIV-oxo intermediate. *Biochemistry* **44**, 3845–3855 (2005).
15. Elkins, J. M. et al. Structure of Factor-inhibiting Hypoxia-inducible Factor (HIF) reveals mechanism of oxidative modification of HIF-1. *J. Biol. Chem.* **278**, 1802–1806 (2003).
16. Liang, G., Klose, R. J., Gardner, K. E. & Zhang, Y. Yeast Jhd2p is a histone H3 Lys4 trimethyl demethylase. *Nature Struct. Mol. Biol.* **14**, 243–245 (2007).
17. Seward, D. J. et al. Demethylation of trimethylated histone H3 Lys4 *in vivo* by JARID1 JmjC proteins. *Nature Struct. Mol. Biol.* **14**, 240–242 (2007).

METHODS

Protein expression and purification. A construct encoding JMJD2A residues 1–359 was amplified by PCR from expressed sequence tag (EST) clone (IMAGE3138875) and cloned into pNIC28-Bsa4. The expression plasmid was transformed into a phage-resistant derivative of *E. coli* BL21(DE3) carrying the pRARE2 plasmid. Twelve litres of TB media containing $50\,\mu g\,ml^{-1}$ kanamycin and $34\,\mu g\,ml^{-1}$ chloramphenicol were inoculated with $5\,ml\,l^{-1}$ overnight culture and grown at $37\,^{\circ}C$ in 2.5-l baffled flasks. Expression was induced with 0.2 mM IPTG at $A_{600} = 0.8$ and allowed to continue for 18 h at $18\,^{\circ}C$. The cells were collected by centrifugation and frozen at $-80\,^{\circ}C$.

Frozen cell pellets were thawed and re-suspended in a total volume of 400 ml lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, 0.5 mM PMSF and $15\,units\,ml^{-1}$ benzonase). Cells were disrupted by high-pressure homogenization followed by sonication and cell debris was removed by centrifugation. Recombinant JMJD2A was purified by Ni-affinity (1 ml HisTrap FF crude) and size-exclusion chromatography (120 ml HiLoad 16/60 Superdex 200) on an AKTAxpress system. The supernatant was loaded onto the Ni-affinity column at $0.8\,ml\,min^{-1}$, washed with 10 column volumes of lysis buffer, 50 column volumes of wash buffer (lysis buffer with 40 mM imidazole), and then eluted with buffer containing 250 mM imidazole. The A280 peak was automatically collected and loaded on the gel filtration column at $1.0\,ml\,min^{-1}$ in 10 mM HEPES, pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP. Pooled fractions had EDTA added to a final concentration of 1 mM and were left overnight. The protein was further purified using an anion exchange column (5 ml HiTrap Q HP) developed with a gradient from 50 to 500 mM NaCl in 50 mM Tris-HCl pH 8.5 (all columns supplied by GE Healthcare). Finally, the protein was exchanged into 10 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP, concentrated to $11\,mg\,ml^{-1}$, flash frozen in liquid N_2 and stored at $-80\,^{\circ}C$.

Crystallization, data collection and model quality. Crystals were grown by vapour diffusion at $4\,^{\circ}C$ by mixing protein solution (100 nl) and well solution (50 nl) and equilibrating versus a reservoir containing 0.1 M citrate pH 5.5, 20% PEG 3350, 4 mM $NiCl_2$. Crystals were transferred to well solution supplemented with 10 mM NOG, 30 mM peptide and 25% glycerol and flash-cooled in liquid N_2. All models were produced from data collected on a single crystal at 100 K. Data sets for the JMJD2A complexes with Ni(II)–NOG and with Ni(II)–NOG–H3K9me3 were collected on beamline SLS-X10SA at a wavelength of 1 Å. The Ni(II)–NOG–H3K9me2 complex data was collected on SLS-X10SA at 0.982 Å. The Ni(II)–NOG–H3K9me1 complex data was collected on ESRF-ID23-1 at 1.072 Å. The Ni(II)–NOG–H3K36me3 complex data was collected using a Rigaku FR-E Superbright generator at 1.542 Å. Data sets were processed with MOSFLM and SCALA. Iterative rounds of model building in COOT and refinement using REFMAC5 gave the final models. All residues were in acceptable regions of a Ramachandran plot as calculated by MolProbity.

Peptide synthesis. Peptides H3K9me3 ARKme3STGGK, H3K9me2 ARKme2STGGK, H3K36me3 GVKme3KPHRY, H3K36me2 GVKme2KPHRY and H3K27me3 mutants AARKme3SAGAT and AARKme3SPAAT were prepared using a CS-Bio CS336S machine, using PL-AMS resin (Polymer Laboratories) and a Rink amide linker. Fmoc-protected amino acids were coupled with DIC/HOAt. Cleavage of the completed peptides from the resin used TFA/triethylsilane (97.5/2.5). Peptides were purified by HPLC (C18 column); predicted masses were confirmed by ESI-MS. Other peptides were obtained from commercial sources as indicated in Supplementary Table 1.

MALDI–TOF analysis of enzymatic activity. In typical assays, purified JMJD2A (1 μM) was incubated with 10 μM peptide in 50 mM Tris-HCl, pH 7.5, 10 μM $(NH_4)_2SO_4 \cdot FeSO_4 \cdot 6H_2O$, 50 μM 2-oxoglutarate, and 2 mM ascorbate (20 min, $37\,^{\circ}C$). Ten microlitres of the demethylation reaction was desalted through a C18 OMIX tip (Varian) according to the manufacturer's instructions. The bound peptides were eluted with $10\,mg\,ml^{-1}$ α-cyano-4-hydroxycinnamic acid MALDI matrix in 60% acetonitrile/0.1% TFA and spotted directly on to the target. Samples were analysed on a MALDI–TOF micro MX mass spectrometer (Waters Micromass) or on an Ultraflex III MALDI TOF–TOF (Bruker Daltonics). Demethylation activity was evaluated for wild-type and modified H3 peptides to investigate the importance of Ser 10, Thr 11, Pro 38 and the Gly 12–Gly 13 motif. H3K27 mutants Pro30Gly and Ala29Pro–Pro30Ala were analysed to test the effect of introducing the H3K9 Gly and H3K36 Pro residue positions into the H3K27 sequence, respectively.

Molecular dynamics studies. Atomic charges for the metals and ligands were obtained using the RESP procedure with electrostatic potential calculated at the B3LYP/lacv3p/IEFPCM ($r = 1.4$ Å, $e = 4$) level of theory. These calculations were performed with Amber 8 and Gaussian 03 suites of programs. Force constants for bonds and angles involving Ni(II) were obtained from the hessian matrix, calculated at the B3LYP/lacvp level for the $Ni(Im)_2(CH_3COO)(CH_3\text{-}NH\text{-}CO\text{-}COO)$ model system. The algorithm fully invariant to the choice of

internal coordinates, suggested by Seminario, implemented in the XYZ-viewer program, was used to extract the force constants. The reference bond lengths and angles were taken directly from the optimized structure of the model system. The van der Waals parameters for nickel were taken from the UFF force field. Parameters for Fe(II) and Zn(II) centres were obtained in an analogous way. For organic cofactors and non-standard amino acid side chains the general AMBER force field (gaff) was used. Atomic charges were determined using the RESP procedure using the electrostatic potential calculated at the B3LYP/cc-pVTZ/IEFPCM ($r = 1.4$ Å, $e = 4$) level.

The standard ionization state for acidic and basic residues was assumed (that is, negatively charged Asp and Glu; positively charged Lys and Arg; and neutral His side chains). The total charge of the system is $+6$. All MD simulations were performed by using the AMBER 8 suite of programs with the FF03 protein force field and the TIP3P water model. The initial structure was solvated in a truncated octahedron of waters with at least 10 Å of water around every atom of the solute, and the positive charge was compensated with $6\,Cl^-$ ions. A standard protocol of minimization and equilibration was used, followed by a 2-ns constant temperature and pressure (300 K, 1 bar) production simulation.

LETTERS

Structural snapshots along the reaction pathway of ferredoxin–thioredoxin reductase

Shaodong Dai[1], Rosmarie Friemann[2]†, Dominique A. Glauser[3]†, Florence Bourquin[3]†, Wanda Manieri[3]†, Peter Schürmann[3] & Hans Eklund[2]

Oxygen-evolving photosynthetic organisms regulate carbon metabolism through a light-dependent redox signalling pathway[1]. Electrons are shuttled from photosystem I by means of ferredoxin (Fdx) to ferredoxin–thioredoxin reductase (FTR), which catalyses the two-electron-reduction of chloroplast thioredoxins (Trxs). These modify target enzyme activities by reduction, regulating carbon flow[2]. FTR is unique in its use of a [4Fe–4S] cluster and a proximal disulphide bridge in the conversion of a light signal into a thiol signal[2]. We determined the structures of FTR in both its one- and its two-electron-reduced intermediate states and of four complexes in the pathway, including the ternary Fdx–FTR–Trx complex. Here we show that, in the first complex (Fdx–FTR) of the pathway, the Fdx [2Fe–2S] cluster is positioned suitably for electron transfer to the FTR [4Fe–4S] centre. After the transfer of one electron, an intermediate is formed in which one sulphur atom of the FTR active site is free to attack a disulphide bridge in Trx and the other sulphur atom forms a fifth ligand for an iron atom in the FTR [4Fe–4S] centre—a unique structure in biology. Fdx then delivers a second electron that cleaves the FTR–Trx heterodisulphide bond, which occurs in the Fdx–FTR–Trx complex. In this structure, the redox centres of the three proteins are aligned to maximize the efficiency of electron transfer from the Fdx [2Fe–2S] cluster to the active-site disulphide of Trxs. These results provide a structural framework for understanding the mechanism of disulphide reduction by an iron–sulphur enzyme[3] and describe previously unknown interaction networks for both Fdx and Trx (refs 4–6).

Fdx provides electrons for regulatory purposes to transmit a light signal from the thylakoid membranes in plants to target enzymes by means of the FTR system (Supplementary Fig. 1)[1]. The FTR cascade begins with the formation of a transient complex between Fdx and FTR (Fig. 1a) and is followed by a straightforward electron transfer from reduced Fdx to FTR (see below; species 1 → 2).

FTR, found only in oxygenic photosynthetic cells, is a thin, flat, heterodimeric molecule. It contains a catalytic and a variable subunit, and has a width of about 10 Å in the centre of the catalytic subunit where the [4Fe–4S] cluster is located[7]. On one side of the disk-shaped molecule, the redox-active disulphide covers the iron–sulphur centre; on the opposite side, a Cys-cis-Pro-Cys (CPC) motif contributes both cysteines to ligate the iron (Fig. 2a). In the structure of the Fdx–FTR complex, Fdx docks at the side of FTR that contains the CPC motif and interacts only with the catalytic subunit. No significant conformational changes in either FTR or Fdx occur when they bind
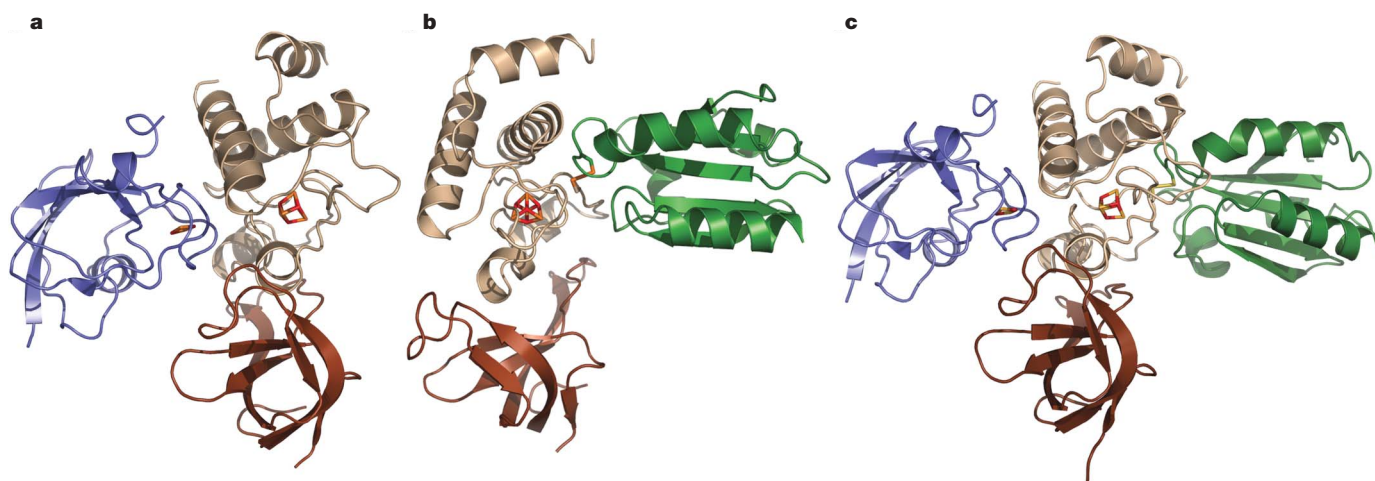

Figure 1 | Overall structures of Fdx–FTR, FTR–Trx-_f_(C49S) and Fdx–FTR–Trx-_f_(C49S) complexes. Fdx, the catalytic subunit of FTR, the variable subunit of FTR and Trx-_f_(C49S) are coloured blue, beige, brown and green, respectively. The [2Fe–2S] cluster of Fdx and [4Fe–4S] cluster of FTR are represented in sticks, where iron and sulphur atoms are coloured red and orange, respectively. **a**, Overall structure of the Fdx–FTR complex. Fdx interacts solely with the catalytic subunit of FTR. **b**, Crystal structure of FTR–Trx-_f_(C49S) complex. Trx-_f_ exclusively interacts with the catalytic subunit of FTR. **c**, Structure of the Fdx–FTR–Trx-_f_(C49S) complex.

[1]Howard Hughes Medical Institute, Integrated Department of Immunology, National Jewish Medical and Research Center & University of Colorado Health Sciences Center, 1400 Jackson Street, Denver, Colorado 80206, USA. [2]Department of Molecular Biology, Swedish University of Agricultural Sciences, Biomedical Centre, Box 590, S-75124 Uppsala, Sweden. [3]Université de Neuchâtel, Laboratoire de Biologie Moléculaire et Cellulaire, Rue Emile Argand 11, CH-2009 Neuchâtel, Switzerland. †Present addresses: Department of Chemistry, University of Oslo, PB 1033, Blindern, 0315 Oslo, Norway (R.F.); Fondation pour Recherches Médicales, CH-1211 Genève, Switzerland (D.A.G.); Biochemisches Institut, Universität Zürich, CH-8057 Zürich, Switzerland (F.B.); College of Pharmacy, Medical Center, University of Cincinnati, Cincinnati, Ohio 45267-0004, USA (W.M.).

92

each other[7–9]. The shortest distance between the edges of the iron–sulphur centres of the two proteins is about 11 Å. Although electron transfer through space would, at this distance, be possible, it would be appreciably faster through a hydrogen- and covalent-bonded pathway[10]. Such a pathway probably exists, because Ser 38 of Fdx forms a hydrogen bond to the carbonyl oxygen of Cys 74 in the FTR CPC motif, providing a short connection between the [2Fe–2S] cluster of Fdx and the [4Fe–4S] cluster of FTR (Supplementary Fig. 2).

Figure 2 | Interactions of Fdx–FTR, FTR–Trx-*f*(C49S) and Fdx–FTR–Trx-*f*(C49S) complexes. a, b, Interactions between the catalytic subunit of FTR and **a**, Fdx and **b**, Trx-*f*(C49S). A disulphide bridge between Cys 57 of FTR and Cys 46 of Trx-*f*, forming the reaction intermediate, links the active sites of FTR and Trx. **c,** The close-up view of the active sites of the Fdx–FTR–Trx-*f*(C49S) complex. Electron transfer from the [2Fe–2S] cluster of Fdx to the intermolecular disulphide bond between the catalytic subunit of FTR and Trx-*f*(C49S) by way of the [4Fe–4S] cluster of FTR is represented as a red broken line. Hydrogen bonds and salt bridges are represented as orange and blue broken lines, respectively. Labels for amino acids in FTR are shown upright, and those for Fdx and Trx are shown in italics.

Although several positively charged residues in FTR and several negatively charged residues of Fdx were expected to be involved in the Fdx–FTR interaction[11], only one such interaction occurs: between FTR Lys 47 and Fdx Glu 92 (Fig. 2a). The charged surfaces thus serve mainly as general attractants rather than providing specificity. The variable subunit of FTR is not involved in the interaction; however, it may act as an additional attractant to facilitate the binding of Fdx to FTR. On the Fdx-binding side of the variable subunit there is a negatively charged patch that is complementary to a positively charged patch of Fdx. In photosynthetic cells, Fdx provides reducing equivalents to Fdx-NADP$^+$-reductase (FNR) to produce NADPH[12]. Interestingly, Fdx binds FTR and FNR in totally different manners (Supplementary Fig. 3).

After Fdx has delivered its electron to FTR, formation of the FTR intermediate **2** (which has been reduced by the addition of one electron, that is, 'one-electron-reduced') in the reaction cycle results from a direct transfer of the first electron through the unique iron of the cluster to the sulphur atom of Cys 87. Together with another electron, provided by the unique iron, the cluster is formally oxidized to the +3 oxidation state and the disulphide is cleaved[11,13]. Further insights into the mode of action of FTR have come from the *N*-ethylmaleimide modified enzyme (NEM-FTR). This chemically modified enzyme, in which the accessible active-site Cys has been alkylated, represents a stabilized form of a one-electron-reduced reaction intermediate (similar to species **2** in Fig. 3)[14]. The electron density map of this one-electron-reduced intermediate unambiguously shows that Cys 87 is coordinated to the unique iron atom (Fig. 4c), different from FTR at resting state, where Cys 87 is in van der Waals contact with the iron (species **1**) (Fig. 4a). This is in agreement with spectroscopic studies that demonstrated that the redox cycle of FTR involves interactions of the [4Fe–4S]$^{3+}$ cluster by means of a fifth cysteinate, probably on the unique iron[13,15]. Multi-coordinated geometry for a unique iron of a [4Fe–4S] cluster has been observed in proteins such as aconitase and radical *S*-adenosylmethionine proteins, in which they are coordinated by either oxygen or nitrogen[16–20]. However, penta-coordination involving two cysteine ligands at a unique iron site, as in the NEM-FTR cluster, has not been observed before.

The involvement of the cluster in forming an intermediate gives it a role beyond a simple electron transfer from the [4Fe–4S] cluster to the active-site disulphide. The [4Fe–4S]$^{2+/1+}$ and [4Fe–4S]$^{2+/3+}$ redox couples of FTR differ substantially from those of Fdx and Trx (refs 21, 22). However, by forming a five-coordinated intermediate, the redox potential of the [4Fe–4S]$^{2+/3+}$ couple of NEM-FTR is lowered from +420 mV to –210 mV, close to the redox potential of the Trx active-site disulphide.

The unique catalytic iron of one-electron-reduced FTR adopts distorted trigonal bipyramidal geometry. The distance between Cys 87 and the unique iron of the centre is reduced from 3.1 Å in the oxidized FTR to 2.6 Å as it becomes a thiolate ligand to the iron ion[23]. This long iron–sulphur bond has a lower energy than a normal bond of this type, which explains why it is readily cleaved when the one-electron-reduced FTR accepts an electron from Fdx. The Cys–iron distances, which influence the redox potential in different complexes, are given in Supplementary Table 2.

In the one-electron-reduced FTR (species **2**), the disulphide bridge is broken and the side chain of the more exposed active-site cysteine, Cys 57, is suited for a nucleophilic attack on the disulphide of Trx, forming a transient intermolecular disulphide bond (species **3**). Two types of Trx (Trx-*f* and Trx-*m*, which have different target enzyme specificities and different phylogenetic origins) in chloroplasts of higher plants use the same redox chemistry to regulate different target enzymes[24]. As predicted, a covalent bond is formed in the complex structures between the exposed Cys of Trx (Cys 46 and Cys 37 of Trx-*f* and Trx-*m*, respectively) and Cys57 of the catalytic subunit of FTR on the side opposite to the Fdx-binding site (Fig. 1b and Supplementary Fig. 4a). The use of Trx-mutants (Trx-*f*(C49S)

and Trx-m(C40S)) prevents scission of this intermolecular disulphide bond by the second Cys of the Trx active site.

The FTR–Trx complexes, which represent one-electron-reduced reaction intermediates containing an oxidized cluster in the +3 oxidation state (**3**), are stabilized by a number of interactions in addition to the intermolecular disulphide. The main interactions (Fig. 2b and Supplementary Fig. 4b) are between residues 82 to 90 of Trx-f (73 to 81 of Trx-m) and residues 57 to 61 of the catalytic subunit of FTR. These are completely conserved among FTRs, suggesting their crucial role in the function of FTR. Only local conformational changes occur on complex formation. The largest differences are caused by changes in the side-chain conformations at the Trx interaction face, which may contribute to the different specificities for Trx-f and Trx-m. Like Fdxs, the Trxs interact exclusively with the catalytic subunit of FTR. The interaction of Trx-f with FTR is similar to that of Trx-m, and the conformational difference between the complexes is a small rotation of the Trx molecule perpendicular to the flat FTR molecule (Supplementary Fig. 5).

The conserved Trp in the WCGPC motif of the Trxs provides a flat surface for interaction with its partners, mainly with hydrophobic residues[6]. In these complexes, Trx-f Trp 45 and Trx-m Trp 36 are not positioned to cover the surface of the Trx molecule in the common Trx conformation, but are instead flipped out, making interactions with FTR (Fig. 2b and Supplementary Fig. 4b). The interactions we observe differ substantially from those between the flavoenzyme Trx-reductase and Trx (Supplementary Fig. 6). The FTR–Trx complexes also demonstrate the presence of a five-coordinate iron atom similar

to that in NEM-FTR (Figs 2b, 4d and Supplementary Fig. 4b). This confirmed that the one-electron-reduced FTR indeed is a reaction intermediate in which the disulphide bond has been cleaved and the exposed reactive FTR Cys 57 has been freed to react with Trx.

FTR–Trx complexes can be efficiently reduced by Fdxs. This requires the formation of a transient Fdx–FTR–Trx complex (**3**→**5** in Fig. 3). Experimental evidence for such complexes has been obtained by chromatography and difference spectroscopy[25]. The crystal structure of this ternary complex shows that FTR accommodates Fdx and Trx-f simultaneously without major conformational changes (Fig. 1c, 2c and Supplementary Fig. 9) because the overall structure of the Fdx–FTR–Trx-f complex is similar to those of the FTR–Trx-f, FTR–Trx-m and Fdx–FTR complexes (Supplementary Fig. 7). However, the molecules are rotated in relation to each other, similar to the different position of Trx-f and Trx-m in relation to FTR (Supplementary Fig. 5). There is an extra salt bridge interaction between FTR Glu 61 and Trx-f Arg 87 in this ternary complex, which is absent in FTR–Trx-f complex. The buried surface area between Fdx and FTR increases from 430 Å^2 in the Fdx–FTR complex to 680 Å^2 in the ternary complex, whereas the interface area between FTR and Trx-f increases from 650 Å^2 in the FTR–Trx-f complex to 830 Å^2, indicating that the ternary complex is more stable.

The unusually thin FTR is ideal for electron transfer and for allowing fast access for the delivery of electrons from Fdx to the disulphide on the other side of the molecule. The shortest distance between the iron centre of Fdx and a cysteine of Trx is approximately 20 Å. The [2Fe–2S], [4Fe–4S] and redox-active disulphide bonds of both FTR
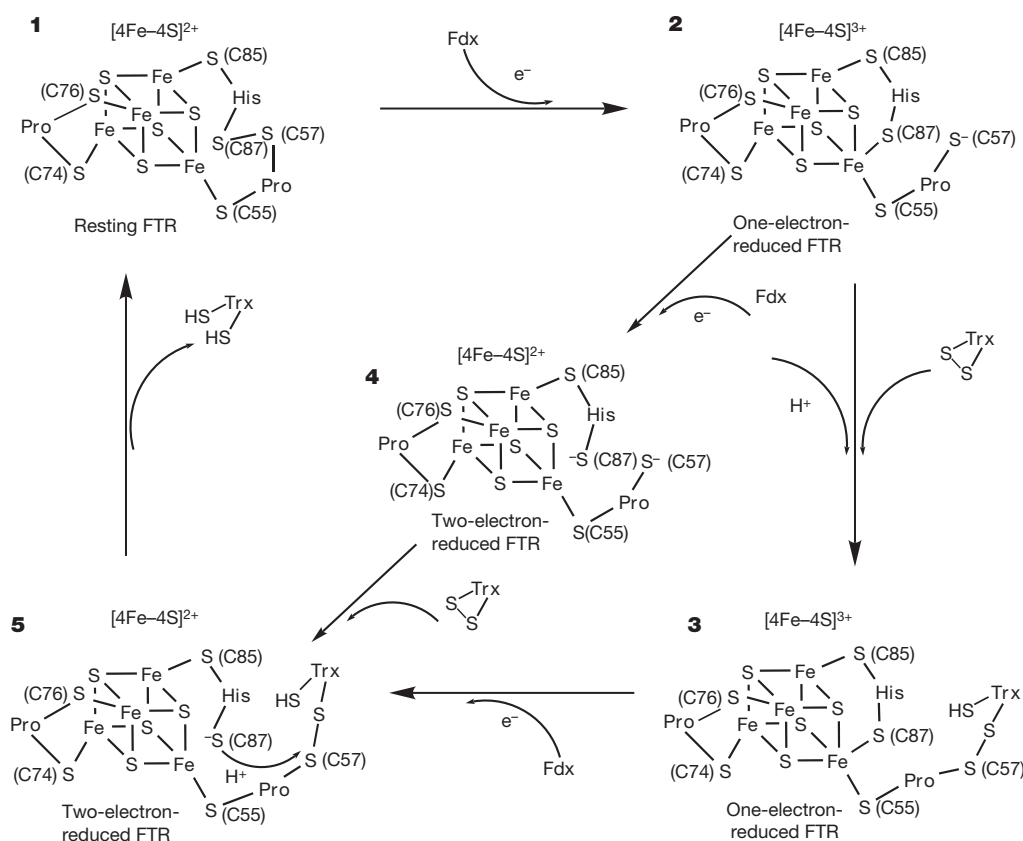

Figure 3 | FTR mechanism proposed on the basis of current structural and spectroscopic[15] investigations. One electron from Fdx, together with one electron from the iron–sulphur cluster, is used to cleave the disulphide bridge of FTR. Cys 87 is stabilized by forming the fifth ligand to the cluster, and Cys 57 is free to attack the disulphide bond of Trx. The cluster of this species **2** shows +3 oxidation state. After thiol–disulphide interchange between Cys 57 of FTR and Trx, the disulphide bond of Trx is cleaved and a FTR–Trx complex with a mixed disulphide bond is formed as species **3**. The second electron can be delivered by a new Fdx either to the

one-electron-reduced FTR or to the FTR–Trx complex. Here, the two-electron-reduced species **4** can form and the reaction pathway splits. In both cases, Fdx reduces the iron–sulphur cluster back to its original +2 oxidation state and Cys 87 is freed. Then Cys 87 attacks the heterodisulphide bridge of FTR–Trx complex (short-lived species **5**), releasing the fully reduced thioredoxin, and the FTR active-site disulphide bond reforms (species **1**). The two-electron-reduced FTR species **4**, with two free thiols, can instantly reduce and release the reduced Trx.

and Trx are aligned in a straight line—an alignment that should maximize the electron-transfer rate (Fig. 2c). Here the complex structure suggests an electron transfer pathway from a [2Fe–2S] cluster to the substrate disulphide bridge that involves several reactions: electron transfer from [2Fe–2S] to [4Fe–4S] clusters, cleavage of a disulphide bond by the [4Fe–4S] centre, and dithiol–disulphide interchange.

Before its interaction with Trx, the one-electron-reduced FTR intermediate can accept a second electron and yield a two-electron-reduced species (4 in Fig. 3); this may lead to an alternative reaction pathway, as suggested in recent spectroscopic studies[15]. In the structure of two-electron-reduced FTR, the disulphide bridge is broken, FTR Cys 87 is detached from the unique iron, and the substrate-interacting FTR Cys 57 is exposed and hydrogen-bonded to a water molecule (Fig. 4b). FTR His 86 rotates towards the cluster and interacts with sulphur atoms of the $[4Fe–4S]^{2+}$ cluster and with FTR Cys 55 (Supplementary Fig. 8). The hydrogen-bonding interactions involving His 86 and Cys 55 or the proximity of the positively charged His 86 to the cluster may be responsible for the anomalous electronic properties of the $[4Fe–4S]^{2+}$ cluster[15,26]. The proximity of His 86 to the iron–sulphur centre can also increase the redox potential of FTR[27], facilitating the reduction of the one-electron-reduced intermediates.

Because one-electron-reduced (species 2) and two-electron-reduced (species 4) FTR should both be able to cleave the disulphide bridge with the reactive Cys 57 and form the intermolecular heterodisulphide bond intermediate, the FTR reduction efficiency may be increased. The two-electron-reduced species can interact with oxidized Trx and form a FTR–Trx complex (species 5) in a mechanism analogous to the one-electron-reduced intermediate. However, this complex should be rather short-lived, because the freed Cys 87 thiol is close to Cys 57 and ready to attack and cleave the heterodisulphide to complete the reduction of Trx and restore the active-site disulphide of FTR. Under normal conditions, the one-electron pathway should be preferred, but, under over-reducing conditions, which occur when there is not enough electron acceptor (for example, CO_2), the two-electron pathway could become more important.

We have captured FTR in all stable redox states in a catalytic cycle. The results illustrate how a disulphide bridge can be cleaved with reducing power donated by an iron–sulphur centre. Our structural analysis shows that only the substrate-exposed unique iron of the cluster performs the chemistry and that coordination variation plays a pivotal role in the catalytic function of the cluster. These results provide key insights into the important site-specific chemistry of the iron–sulphur cluster and reveal its novel mode of action. They may also have implications for the mechanisms of related [4Fe–4S]-containing proteins, such as the radical S-adenosylmethionine enzymes and heterodisulphide reductases[26,28,29].

METHODS SUMMARY

Synechocystis FTR and Fdx, mutant spinach Trx-f and Trx-m were overexpressed in *Escherichia coli*. NEM-FTR was obtained following the method described earlier[30]. Two-electron-reduced FTR was obtained by reducing the crystals of oxidized FTR (ref. 11) with 100 mM dithionite in presence of 50–100 mM methyl viologen for 2 h using gradient soaking in anaerobic conditions. The non-covalent Fdx–FTR complex was obtained by mixing the two proteins at equimolar concentrations and incubating overnight at 4 °C. Covalent FTR−Trx complexes were obtained by mixing wild-type FTR with an equimolar amount of Trx-mutant. The mixture was reduced with 10 mM dithiothreitol (DTT), diafiltered with the same buffer without DTT and left overnight at 4 °C under aerobic conditions. Fdx–FTR–Trx-f complex was obtained by mixing Fdx and the FTR–Trx-f complex at equimolar concentrations. All proteins were crystallized by hanging- or sitting-drop vapour diffusion methods. X-ray diffraction data were collected under liquid-nitrogen cryoconditions at 100 K. Data were processed with the HKL2000 or HKL suite of programs (HKL Research). The crystal structures were determined by the molecular replacement method. Oxidized FTR was used as the search model for NEM−FTR and reduced FTR. Oxidized FTR, Trx or Fdx were used for the molecular replacement calculation of the complexes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Buchanan, B. B. & Balmer, Y. Redox regulation: a broadening horizon. *Annu. Rev. Plant Biol.* 56, 187–220 (2005).
2. Schürmann, P. Redox signaling in the chloroplast: the ferredoxin/thioredoxin system. *Antioxid. Redox Signal.* 5, 69–78 (2003).
3. Johnson, D. C., Dean, D. R., Smith, A. D. & Johnson, M. K. Structure, function, and formation of biological iron–sulfur clusters. *Annu. Rev. Biochem.* 74, 247–281 (2005).
4. Kurisu, G. et al. Structure of the electron transfer complex between ferredoxin and ferredoxin-NADP⁺ reductase. *Nature Struct. Biol.* 8, 117–121 (2001).
5. Morales, R. et al. A redox-dependent interaction between two electron-transfer partners involved in photosynthesis. *EMBO Rep.* 1, 271–276 (2000).
6. Lennon, B. W., Williams, C. H. Jr & Ludwig, M. L. Twists in catalysis: alternating conformations of *Escherichia coli* thioredoxin reductase. *Science* 289, 1190–1194 (2000).
7. Dai, S. et al. How does light regulate chloroplast enzymes? Structure–function studies of the ferredoxin/thioredoxin system. *Q. Rev. Biophys.* 33, 67–108 (2000).
8. van den Heuvel, R. H. et al. The active conformation of glutamate synthase and its binding to ferredoxin. *J. Mol. Biol.* 330, 113–128 (2003).
9. Xu, X. et al. Ferredoxin/ferredoxin–thioredoxin reductase complex: complete NMR mapping of the interaction site on ferredoxin by gallium substitution. *FEBS Lett.* 580, 6714–6720 (2006).
10. Onuchic, J. N., Beratan, D. N., Winkler, J. R. & Gray, H. B. Pathway analysis of protein electron-transfer reactions. *Annu. Rev. Biophys. Biomol. Struct.* 21, 349–377 (1992).
11. Dai, S., Schwendtmayer, C., Schürmann, P., Ramaswamy, S. & Eklund, H. Redox signaling in chloroplasts: cleavage of disulfides by an iron–sulfur cluster. *Science* 287, 655–658 (2000).
12. Knaff, D. B. in *Advances in Photosynthesis* (eds Ort, D. R. & Yocum, C. F.) 333–361 (Kluwer Academic, Dordrecht, 1996).
13. Jameson, G. N. et al. Spectroscopic evidence for site specific chemistry at a unique iron site of the [4Fe–4S] cluster in ferredoxin:thioredoxin reductase. *J. Am. Chem. Soc.* 125, 1146–1147 (2003).
14. Staples, C. R. et al. Role of the [Fe₄S₄] cluster in mediating disulfide reduction in spinach ferredoxin:thioredoxin reductase. *Biochemistry* 37, 4612–4620 (1998).
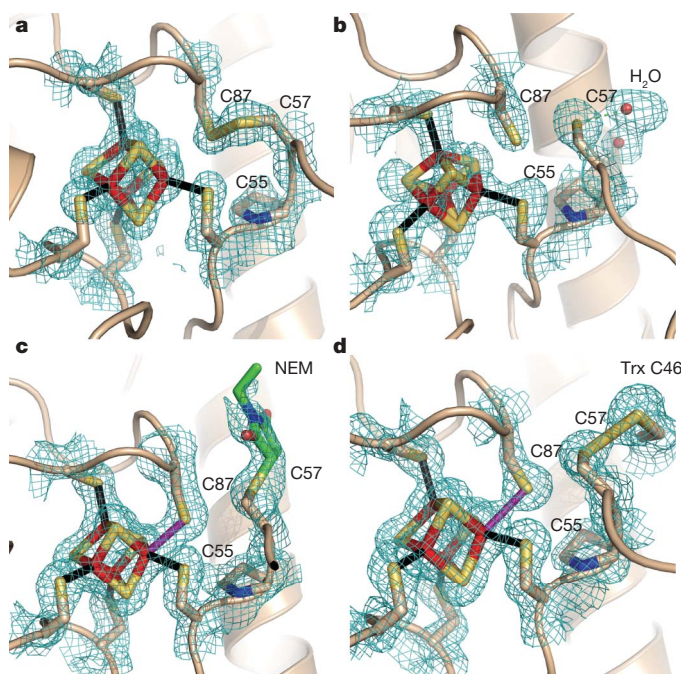

Figure 4 | Comparison of the active sites of FTR at different reaction states. Simulated annealing $2F_o–F_c$ omit electron maps, in which FTR Cys 55, Cys 57, Cys 87 and the [4Fe–4S] cluster were omitted, contoured at 1σ around the active-site disulphide and [4Fe–4S] cluster of FTR. The [4Fe–4S] cluster of FTR is represented as sticks, in which iron and sulphur atoms are coloured red and orange, respectively. Water molecules are shown as red spheres. **a**, Resting enzyme[11]. **b**, Two-electron-reduced FTR. **c**, NEM-FTR. **d**, FTR–Trx-f(C49S) complex.

15. Walters, E. M. et al. Spectroscopic characterization of site-specific [Fe₄S₄] cluster chemistry in ferredoxin:thioredoxin reductase: implications for the catalytic mechanism. *J. Am. Chem. Soc.* **127,** 9612–9624 (2005).

16. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. Crystal structure of biotin synthase, an *S*-adenosylmethionine-dependent radical enzyme. *Science* **303,** 76–79 (2004).

17. Hanzelmann, P. & Schindelin, H. Crystal structure of the *S*-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. *Proc. Natl Acad. Sci. USA* **101,** 12870–12875 (2004).

18. Lauble, H., Kennedy, M. C., Beinert, H. & Stout, C. D. Crystal structures of aconitase with trans-aconitate and nitrocitrate bound. *J. Mol. Biol.* **237,** 437–451 (1994).

19. Layer, G., Moser, J., Heinz, D. W., Jahn, D. & Schubert, W. D. Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes. *EMBO J.* **22,** 6214–6224 (2003).

20. Lepore, B. W., Ruzicka, F. J., Frey, P. A. & Ringe, D. The X-ray crystal structure of lysine-2,3-aminomutase from *Clostridium subterminale. Proc. Natl Acad. Sci. USA* **102,** 13819–13824 (2005).

21. Staples, C. R. et al. The function and properties of the iron–sulfur center in spinach ferredoxin:thioredoxin reductase: a new biological role for iron–sulfur clusters. *Biochemistry* **35,** 11425–11434 (1996).

22. Hirasawa, M. et al. Oxidation–reduction properties of chloroplast thioredoxins, ferredoxin:thioredoxin reductase, and thioredoxin *f*-regulated enzymes. *Biochemistry* **38,** 5200–5205 (1999).

23. Giastas, P. et al. The structure of the 2[4Fe–4S] ferredoxin from *Pseudomonas aeruginosa* at 1.32-Å resolution: comparison with other high-resolution structures of ferredoxins and contributing structural features to reduction potential values. *J. Biol. Inorg. Chem.* **11,** 445–458 (2006).

24. Buchanan, B. B., Schürmann, P., Decottignies, P. & Lozano, R. M. Thioredoxin: a multifunctional regulatory protein with a bright future in technology and medicine. *Arch. Biochem. Biophys.* **314,** 257–260 (1994).

25. Glauser, D. A., Bourquin, F., Manieri, W. & Schürmann, P. Characterization of ferredoxin:thioredoxin reductase modified by site-directed mutagenesis. *J. Biol. Chem.* **279,** 16662–16669 (2004).

26. Walters, E. M. & Johnson, M. K. Ferredoxin:thioredoxin reductase: disulfide reduction catalyzed via novel site-specific [4Fe–4S] cluster chemistry. *Photosynth. Res.* **79,** 249–264 (2004).

27. Chen, K. et al. Crystal structures of ferredoxin variants exhibiting large changes in [Fe–S] reduction potential. *Nature Struct. Biol.* **9,** 188–192 (2002).

28. Duin, E. C., Madadi-Kahkesh, S., Hedderich, R., Clay, M. D. & Johnson, M. K. Heterodisulfide reductase from *Methanothermobacter marburgensis* contains an active-site [4Fe–4S] cluster that is directly involved in mediating heterodisulfide reduction. *FEBS Lett.* **512,** 263–268 (2002).

29. Chen, D., Walsby, C., Hoffman, B. M. & Frey, P. A. Coordination and mechanism of reversible cleavage of *S*-adenosylmethionine by the [4Fe–4S] center in lysine 2,3-aminomutase. *J. Am. Chem. Soc.* **125,** 11788–11789 (2003).

30. Schürmann, P. & Gardet-Salvi, L. Chemical modification of the active site of ferredoxin–thioredoxin reductase. *Chimia* **47,** 245–246 (1993).

METHODS

Sample preparation and crystallization. *Synechocystis* FTR shows no functional difference from spinach FTR, but is significantly more stable[31]. FTR and Fdx from *Synechocystis* sp. PCC6803 and mutant spinach Trx-*f*(C49S) were expressed and purified, as previously described[25,32]. Mutant spinach Trx-*m*(C40S) is based on the recombinant Trx-*m* described earlier[33], from which the amino-terminal extra peptide of six residues was removed to obtain a protein corresponding in its size exactly to spinach Trx-*m*[34]. The mutant construct was cloned into the expression plasmid pET-3c and expressed in *E. coli* strain BL21(DE3).

NEM-FTR was obtained following the method described in detail earlier[30]. FTR was reduced in the light with thylakoids and Fdx under argon using thylakoids capable of only photosystem I (in which there is no oxygen evolution), with ascorbate/dichloro-phenolindophenol as the electron donor system instead of water. After five minutes in the light, NEM (Fluka) was added, and incubation continued for two minutes. Excess NEM was quenched by addition of 2-mercaptoethanol as the light was turned off. The mixture was centrifuged, desalted on G-25 column, and the NEM-FTR separated from Fdx and contaminating proteins from the thylakoids by ion exchange chromatography[30]. NEM-FTR was crystallized at room temperature by hanging-drop vapour diffusion against 1 ml mother liquor containing 2.1 M ammonium sulphate and 100 mM sodium acetate, pH 5.4. In all crystallization setups, 2.0 µl of protein solution (16 mg ml^{-1}, in 20 mM triethanolamine (TEA)-Cl buffer, pH 7.5) was mixed with an equal volume of reservoir solution.

Crystals of oxidized FTR were grown at room temperature by the sitting-drop vapour diffusion method. Each drop, consisting of 2 µl protein solution and an equal volume of reservoir solution, was equilibrated against 1 ml reservoir solution containing 1.8 M ammonium sulphate and 100 mM sodium acetate buffer around pH 5.2 (ref. 11). The sample was concentrated to 24 mg ml^{-1} in 50 mM Tris-Cl buffer pH 7.5. The crystals were usually obtained with the help of microseeding technique[35]. All crystallization trials were performed in an anaerobic chamber with an O_2 level of less than 5 parts per million. The solutions were degassed thoroughly before moving into the glove box. Two-electron-reduced FTR was obtained by reducing the crystals of oxidized FTR with 100 mM dithionite in presence of 50–100 mM methyl viologen[36] for two hours using gradient soaking.

The noncovalent Fdx–FTR complex was obtained by mixing the two proteins at equimolar concentrations and incubating overnight at 4 °C under conditions that were shown to allow complex formation[25]. Crystallization was performed by the hanging- or sitting-drop vapour diffusion method. In all the experiments, 1.5–2 µl protein solution and 1.5–2 µl precipitant solution were equilibrated against 1 ml precipitant solution. Fdx–FTR complex (27 mg ml^{-1}, in 50 mM Tris-Cl buffer, pH 7.5) crystallized at 20 °C in 32% polyethylene glycol (PEG)8000 and 0.1 M cacodylate, pH 5.5. Crystals normally formed within one week.

Mutation of the buried active-site Cys residue of Trx allows one to stabilize the transient intermolecular disulphide bridge, a method that has been used extensively in several studies[6,37]. Using this approach, we generated covalent complexes between FTR and mutant Trx-*m*(C40S) and Trx-*f*(C49S). Covalent FTR−Trx complexes were obtained by mixing concentrated wild-type FTR with an equimolar amount of Trx mutant. The mixture in 20 mM TEA-Cl buffer, pH 7.3, was reduced with 10 mM DTT, diafiltered with the same buffer without DTT, and left overnight at 4 °C with very slow stirring to allow oxygen to function as electron acceptor. The FTR–Trx complexes formed were re-purified by ion exchange chromatography to remove unreacted proteins from the complex[25]. Crystallization experiments were carried out by the hanging-drop vapour diffusion method. In a typical trial, 2 µl protein samples were mixed with an equal amount of mother liquor and then equilibrated against 1 ml reservoir solution. Crystals of FTR–Trx-*f* (18 mg ml^{-1}) were obtained within two weeks in 1.8–2.0 M ammonium sulphate, 0.1 M KNa tartrate and 0.2 M sodium citrate, pH 5.2–6.0, at 20 °C. Crystals of FTR–Trx-*m* complex (20 mg ml^{-1}) were grown at 15 °C in 20% (w/v) PEG3350, 0.2 M sodium fluoride and 0.1 M HEPES-OH, pH 6.0. Crystals were obtained within three months.

Fdx–FTR–Trx-*f* complex was obtained by incubating at 4 °C overnight after mixing equimolar concentrations of Fdx and FTR–Trx-*f*. Subsequently, the complex was purified with a size-exclusion column, Superdex 75 (GE Healthcare), at 4 °C. Crystallization trials were performed by the vapour-diffusion technique using the hanging-drop method. Optimized crystals were obtained at 20 °C in hanging drops by mixing 1 µl of 30 mg ml^{-1} protein in 20 mM TEA-Cl buffer, pH 7.5, and 1 µl reservoir solution containing 2.0 M ammonium sulphate, 0.1 M KNa tartrate and 0.2 M sodium citrate, pH 5.6.

Structure determination. All X-ray diffraction data were collected under liquid nitrogen cryo-conditions at 100 K. Before refinement, an independent set of 5% reflections was set aside for the free *R* (R_{free}) value calculation for all data sets. No sigma cutoffs were used in all refinements. Both conventional *R*-factor (R_{cryst})

and R_{free} (ref. 38) were used to monitor the progress of refinement. The models were subjected to several rounds of alternating simulated annealing/positional refinement in CNS[39] followed by B-factor refinement in CNS or REFMAC[40]. Model building was performed using the program O[41]. Several residues of the amino termini and carboxy termini of both the catalytic subunit and the variable subunit of FTR are disordered in all structures; however, the majority of the residues are well defined except several solvent-exposed side chains. Simulated annealing omit maps were routinely used to remove the model bias. The CNS parameter and topology files of [4Fe–4S] and [2Fe–2S] clusters were generated by Hetero-compound Information Centre — Uppsala (HIC-Up)[42]. The distance between Cys 87 Sγ and Fe was not restrained. However, the distances between Sγ of the cluster-ligating cysteines and iron were restrained to 2.3 Å in initial refinements. After R_{free} dropped below 30%, the bond restraints between Cys Sγ and iron were lifted in refinements. All models have good stereochemistry, as determined by the program PROCHECK[43]. Data collection and refinement statistics including Ramachandran plot statistics are summarized in Supplementary Table 1. Surface areas were calculated with GRASP[44]. Figures were prepared using PyMOL[45].

The Fdx–FTR complex crystals were flash-cooled in liquid nitrogen after a 10–30 s soak in a cryoprotection solution consisting of the reservoir solution with an addition of 27% (w/v) PEG4000. X-ray data were measured at SBC beamline ID-19 at the APS of Argonne National Laboratory. The data were indexed, integrated, scaled and merged using HKL2000 (ref. 46). The structure of Fdx–FTR was determined by molecular replacement using the program AMoRe[47] by using *Synechocystis* Fdx (Protein Data Bank code 1DOX, ref. 48) and *Synechocystis* oxidized FTR (Protein Data Bank code 1DJ7, ref. 11) as search models.

FTR–Trx-*f* crystals were flash-frozen in a cryoprotection solution consisting of the reservoir solution with addition of 25% (v/v) glycerol in liquid nitrogen. Diffraction data were collected on a Rigaku R2U rotating anode using a MSC RaxisIV image plate detector. The four data sets collected on four crystals were indexed, integrated, scaled and merged using HKL2000 (ref. 46). The structure was determined by the molecular replacement program AMoRe[47] using the structure of *Synechocystis* FTR[11] as a search model. After an initial round of refinement, Trx-*f* could be manually modelled into the $2F_o–F_c$ map using O[41].

FTR–Trx-*m* crystals were flash-cooled in liquid nitrogen without addition of cryoprotectant. Diffraction data were collected on beamline ID14-4 at the ESRF, using an ADSC detector. The data were integrated using MOSFLM[49] and scaled using SCALA[50]. The structure was solved with the molecular replacement program MOLREP[51], using the FTR–Trx-*f* complex as a search model. The FTR–Trx-*m* crystals have two complex molecules in the asymmetric unit, and the electron density is well defined for both FTR–Trx-*m* molecules.

NEM-FTR and two-electron-reduced FTR crystals were flash-frozen in liquid nitrogen after a soak lasting 2–10 s in mother liquor solutions containing 20% glycerol. X-ray data of two-electron-reduced FTR were measured at SBC beamline ID-19 at the APS. The data were indexed and integrated using DENZO and reduced using SCALEPACK[46]. The first model of reduced FTR was obtained by rigid body refinement in REFMAC[40] in CCP4 using the structure of the oxidized FTR[11] as a model. There was a huge negative density (more than 5σ) at the active-site disulphide bond in the different Fourier maps, and a small positive at 3σ on surface of the protein, in agreement with the breakage of the disulphide bond. Data sets of NEM-FTR were collected on beamline 8.2.1 at ALS of the Lawrence Berkeley National Laboratory. Data were processed and scaled with HKL2000 (ref. 46). Molecular replacement was performed using the program AMoRe[47], and the oxidized FTR structure[11] without the iron–sulphur cluster was used as a search model. One single high peak was obtained after rotation and translation functions. In the initial maps, we could clearly see the iron–sulphur cluster and there was an extra positive density extended from the density of Cys 57. NEM could be nicely modelled into this positive density. Phase improvement and refinement of both structures were performed by CNS[39], and model was manually adjusted using the program O[41].

Fdx–FTR–Trx-*f* crystals were flash-frozen in a cryoprotection solution consisting of the reservoir solution with addition of 20% (v/v) glycerol in liquid nitrogen. X-ray diffraction data were collected at beamline 8.2.1 at ALS using a CCD detector. The data were indexed, integrated, scaled and merged using HKL2000 (ref. 46). The structure was determined by the molecular replacement program AMoRe[47] using the structures of FTR–Trx-*f* and *Synechocystis* Fdx as search models. The solution was unambiguous and showed presence of one complex within the asymmetric unit.

31. Manieri, W. *et al.* N-terminal truncation of the variable subunit stabilizes spinach ferredoxin:thioredoxin reductase. *FEBS Lett.* **549**, 167–170 (2003).
32. Balmer, Y. & Schürmann, P. Heterodimer formation between thioredoxin *f* and fructose 1,6-bisphosphatase from spinach chloroplasts. *FEBS Lett.* **492**, 58–61 (2001).

33. Schürmann, P. Ferredoxin:thioredoxin system. *Methods Enzymol.* **252,** 274–283 (1995).

34. Wedel, N., Clausmeyer, S., Herrmann, R. G., Gardet-Salvi, L. & Schürmann, P. Nucleotide sequence of cDNAs encoding the entire precursor polypeptide for thioredoxin *m* from spinach chloroplasts. *Plant Mol. Biol.* **18,** 527–533 (1992).

35. Stura, E. A. in *Crystallization of Nucleic Acids and Proteins. A Practical Approach* (eds Ducruix, A. & Giegé, R.) 99–126 (Oxford Univ. Press, New York, 1999).

36. Schürmann, P., Stritt-Etter, A. L. & Li, J. Reduction of ferredoxin:thioredoxin reductase by artificial electron donors. *Photosynth. Res.* **46,** 309–312 (1995).

37. Balmer, Y. *et al.* Proteomics gives insight into the regulatory function of chloroplast thioredoxins. *Proc. Natl Acad. Sci. USA* **100,** 370–375 (2003).

38. Brunger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–475 (1992).

39. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).

40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).

41. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).

42. Kleywegt, G. J., Henrick, K., Dodson, E. J. & van Aalten, D. M. Pound-wise but penny-foolish: how well do micromolecules fare in macromolecular refinement? *Structure* **11,** 1051–1059 (2003).

43. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).

44. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11,** 281–296 (1991).

45. DeLano, W. The PyMOL Molecular Graphics System. ⟨http://www.pymol.org⟩ (2002).

46. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).

47. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50,** 157–163 (1994).

48. Lelong, C., Setif, P., Bottin, H., Andre, F. & Neumann, J. M. ^{1}H and ^{15}N NMR sequential assignment, secondary structure, and tertiary fold of [2Fe–2S] ferredoxin from *Synechocystis* sp. PCC 6803. *Biochemistry* **34,** 14462–14473 (1995).

49. Powell, H. R. The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr. D* **55,** 1690–1695 (1999).

50. Collaborative Computational Project No 4. The *CCP4* suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).

51. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30,** 1022–1025 (1997).

nature jobs

The poor representation of women and minorities in science is an all-too-familiar tale. Despite some degree of progress, the issue has yet to be fully addressed. Often when magazines and journals tackle the subject, they do so from the perspective of equal rights — emphasizing that these under-represented groups deserve to take their rightful place within the scientific enterprise at universities, in industry and in government.

This is, of course, true. But the issue goes well beyond equal rights. As the demographics in, for example, the United States change, it is imperative that women and under-represented minorities have the chance to contribute — not only because they deserve to, but because the scientific enterprise of the future needs their contribution if it is to remain healthy, vibrant and talent-rich. Barriers related to cultural mores, money or child-bearing prevent talented scientists from pursuing research, effectively shrinking the applicant pool. On page 98, we explore some of these barriers in the United States and how they could be addressed.

In Europe, attempts to get more women into science have been complicated by the region's diverse culture and history, as detailed on page 101. But here too, efforts are being made to improve the mobility and career prospects of women scientists. As well as antidiscrimination legislation, there are a number of initiatives such as bids to improve networking and the sharing of career advice among women.

But regardless of whether you're talking about Munich or Maine, one issue is consistent wherever you look: institutions worldwide need to find ways to keep women scientists in the profession beyond the assistant professor stage.

Unfortunately, the upshot of repeated efforts to increase participation of women and minorities all-too-often can be summed up with: "There's been some progress, but much more needs to be done." Nevertheless, it is worth updating researchers and administrators about the state of such efforts. The more that fledgling under-represented scientists know, the better off they will be — and the better off the scientific enterprise will be as well.

Gene Russo, acting editor of *Naturejobs*

Beyond the glass ceiling

Women and under-represented minorities are earning historically high numbers of science doctorates in the United States. So why aren't they making it to the professorial ranks? **Kendall Powell** investigates.

Aaron Velasco describes himself as the only US-born Latino seismologist in the country. As a faculty member at the University of Texas, El Paso, he is part of a rare group of under-represented minorities who make it into tenure-track academic positions in the United States. His story illustrates part of the reason for the abysmally low numbers of others like him.

"I honestly could not afford to become a postdoctoral fellow," says Velasco, recalling how the enormous debt he had built up during almost ten years of studying beyond high school forced him to seek something better than a postdoc's salary. In search of financial security, Velasco went straight into industry after graduate school, then found his way back to academia. Many other excellent minority graduates cite economic disadvantage as a major reason for why they don't end up in academic positions — even though the number of minority PhDs is on the rise.

If academia is to offer varied role models and perspectives for a diverse population of students, it must become more welcoming to women and ethnic minorities, leaders of diversity efforts say. Industry has already learned the value of diversity. In a 2003 amicus brief in support of the University of Michigan's affirmative-action admissions policies, 65 Fortune 500 firms argued that efforts to increase diversity improve innovation, productivity and global competition.

Women and minorities suffer from the effects of isolation once they enter the upper ranks of academia. Both groups perceive academia as an unfriendly environment, and both suffer from an implicit bias against them in the hiring process.

> "I honestly could not afford to become a postdoctoral fellow."
> — Aaron Velasco

For women, the clash of their biological clock with the tenure clock, along with the effort of balancing work and family duties, is a huge barrier to advancing up the academic ladder. For minorities, financial and geographical constraints make academia a less attractive choice than industry. Attempts to remove barriers and to mend holes in the pipeline have met with mixed success.

Leaving academia

In 2003, 51% of the US population was female and more than 25% of the population was from a minority group under-represented in science: African Americans, Latinos and Native Americans. Women earned well over one-third of the science and engineering doctorates awarded in 2003–04 and African American and Latino doctorates have steadily increased during the past ten years (see Tables 1 and 2).

But women hold fewer than one-third of all science and engineering faculty posts, and just 18% of full professorships. For minorities, the numbers are below 10% and 6.7%, respectively. When the numbers are dissected at the disciplinary level, many fields find they are doing far worse in hiring talented women and minorities than should be expected, given the numbers of doctorates they award to those groups (see Table 3). Although many 'diversity in science' programmes have been in place for more than 30 years, the faculty in most US academic science departments has remained overwhelmingly white and male.

The numbers show that not only are women having a hard time reaching parity in the hiring process, but that they continue to struggle for parity at all levels of success such as making tenure, advancing to

administrative positions, and gaining national recognition for scientific achievements. These numbers also send a striking message to the next generation.

"I think young women looking at the PhD-to-faculty transition are being more pragmatic, looking down the road and saying, 'I don't want to beat my head against a wall for the next 20 years,'" says Donna Dean, president of the Association for Women in Science (AWIS) in Washington DC. The AWIS began in 1971 to help women succeed at the mid-career stage. Dean says the focus has shifted to earlier stages, to recognize that women fight an uphill battle from the minute they earn their doctorates.

Women and minorities must both deal with implicit bias, a problem that is well-documented in the social-science literature, but one that has garnered little attention from the science sector until recently. Dean describes the problem of implicit bias in these terms: "People are most comfortable with people who think and look like themselves."

This type of bias cuts across all divides and has been shown to affect everything from basketball refereeing calls to hiring practices. In addition, a strong gender bias has been found in workplace scenarios, with both men and women consistently overrating men and underrating women in job qualifications (see Virginia Valian's chapter in *Why Aren't More Women in Science?* (eds S. J. Ceci and W. M. Williams); American Psychological Association Press, 2006).

Bias cuts

"When you have homogenous, privileged groups it is hard for them to see that their decisions are inhibiting their excellence," says Meg Urry, an astrophysicist and the first woman to chair the physics department at Yale University. Most scientists think they operate in a meritocracy, rewarding excellent research irrespective of colour or gender lines. But the data show that is simply not the case, says Urry. And many scientists, she says, are "unaware of that data and unaware that they have internal biases".

To change that, several groups have begun highlighting research on bias at workshops for different science disciplines. Chemists are leading the way with the help of the Committee on the Advancement of Women Chemists based in Eugene, Oregon, by holding a workshop last year for 55 chairs from the top

Women remain under-represented at the higher end of the academic scale.

chemistry departments around the country (see 'Chemistry case study').

Programmes to recruit and retain university minority students in science have made steady, if small, improvements. The numbers of science bachelor's degrees awarded to minority students, about 16% of the total, is now commensurate with the number of minorities enrolled in university. The number of African American and Latino science doctorates have increased about 20% during the past ten years.

"I'm encouraged by the numbers of kids at the beginning of the pipeline," says Velasco. "But my worry is that these kids will want to go into academia and find their opportunities are limited there."

For minorities, their small numbers mean that feelings of isolation begin early and are likely to persist throughout a career (see 'A political hot potato'). Many under-represented minority students come from disadvantaged backgrounds that make both financial concerns and extended family responsibilities rise to the top of the priority list when they consider a career move.

Isiah Warner, a chemist at Louisiana State University in Baton Rouge, has overseen the 15-year transformation of that department into the top producer of African American chemistry PhDs. But, he

CHEMISTRY CASE STUDY

Chemistry as a field has made some progress towards retaining talented women and minority chemists in the academic ranks. Still, although women gain roughly a third of chemistry doctorates, they hold only 13% of chemistry faculty positions.

In January 2006, the National Science Foundation, the Department of Energy, the National Institutes of Health and chemistry leaders sponsored a workshop in which 55 chairs from the top-ranked departments around the country gathered to face the problem and take action to address it. The workshop highlighted research on implicit bias and on issues affecting women's ability to succeed in academia.

Before the workshop, when participants were asked why women were not being recruited, hired

and retained in their departments, the participants blamed factors largely beyond their control: too few women in the applicant pool, losing females to other departments and no money for recruiting both members of a couple. After the training on implicit bias, participants were more likely to admit to a lack of commitment or downright opposition to hiring female faculty members, says Geraldine Richmond, a chemist at the University of Oregon in Eugene who is evaluating the workshop's impact.

Participants left with a commitment to implement at least two items within their departments or institutions, such as doubling the number of female applicants in the next faculty search, or advocating subsidized childcare. And the participants agreed to evaluate the

effectiveness of their efforts in the future.

Physics and geosciences have followed suit with their own gender-equity workshops. Chemistry leaders are now planning a workshop to address the lack of minority faculty members, with the goal of encouraging departments to cultivate at least one minority faculty candidate in the next five years.

Biological sciences, which have similar gender imbalances, could learn from other disciplines' scientific approach and evaluation of the issue, says Donna Dean of the Association for Women in Science in Washington DC. She notes that the funding agencies for biomedical research have "not stepped up to the plate in paying attention to the changing demographics and what's happening to PhDs as they move into faculty positions".   **K.P.**

A POLITICAL HOT POTATO

One of the obstacles facing minority biomedical scientists could be the way US government funding is distributed through the National Institute of General Medical Sciences (NIGMS) in Bethesda, Maryland. Its Division of Minority Opportunities in Research (MORE) is the largest funder of minority programmes through the National Institutes of Health and has greatly influenced the upward trend of minority PhD numbers. But progress in diversifying faculty has been disappointing or nonexistent, according to Jeremy Berg, director of the NIGMS.

In 2005, MORE was jolted by a working group's report that, although almost 60% of MORE's budget was going to programmes at minority-serving institutions, more than 70% of minority students are receiving their BS degrees from majority institutions. Even though the topic was a "political hot potato", says working group co-chair Virginia Zakian, a molecular biologist at Princeton University in New Jersey, the group recommended that the MORE budget should more closely reflect the realities of where minority students are educated.

Berg, however, says that it would be a mistake to shift MORE money away from minority-serving institutions towards others that have significantly more resources. Instead, he and the NIGMS are considering how best to restructure MORE programmes so that they not only encourage minority students to take PhDs, but also see them through to faculty positions.

Berg says the best way to do that is to make changes to programmes at the institutional level, such as the NIGMS' T32 student training grants which stipulate that the receiving department must have a diversity programme in place. Also, says Berg, there should be partnerships between undergraduate universities and top-tier graduate programmes to ensure that minority PhD students aren't starting off at a disadvantage in the academic career track.

Although Berg and Zakian may disagree on how to get there, they agree wholeheartedly that diversifying the biomedical faculty is vital. "This is not an issue of social justice or equal opportunity," says Berg. "The biomedical workforce is much weaker than we need it to be and student diversity is [outstripping] faculty diversity."   K.P.

notes, financial realities and the unfriendly climate of academia conspire to lure a huge portion of those students directly into industry positions.

"They see me work 8–12 hours a day, seven days a week for a job that pays only two-thirds of your salary, meanwhile you have to hustle the other third of your salary and grant money constantly," says Warner. That's compared with an industry job offering a $90,000 salary out of graduate school and a 40-hour work week. "Which would you choose?" he asks.

Table 1 | Under-represented groups in the science academic pipeline, 2003–04

Group	General population (%)	BS degree (%)	PhDs (%)*	Postdocs (%)	Assistant professor (%)†	Full professor (%)†
Women	51	50	37	40	36	18
African American	13	8	3	3	7	4
Latino	14	7	3	4	4	2
Native American	1	0.7	0.2	0.3	0.8	0.7
White	67	65	46	63	71	85

Source: Commission on Professionals in Science and Technology Professional Women and Minorities *A Total Human Resources Data Compendium* 16th edition (November 2006); www.cpst.org
*US citizens and permanent residents only
†Numbers include faculty members in social sciences, political science and economics, which historically include higher percentages of women and minority PhDs

Table 2 | Science and engineering PhDs granted to under-represented minorities

Group	1996	2005	% change
African American	431	536	24
Latino	529	650	23
Native American	70	48	−31

Source: National Science Foundation/Division of Science Resources Statistics, Survey of Earned Doctorates

Table 3 | Tenure-track faculty in 'Top 50' science departments by discipline, 2001–02

Discipline	Women (%)	African American (%)	Latino (%)	Native American (%)	White (%)
Chemistry	12	1.2	1.8	0.2	90
Physics	6	0.6	1.9	0.1	86
Astronomy	13	1.2	1.2	0.0	91
Mathematics	8	0.9	2.6	0.1	85
Computer Sci.	11	0.3	1.3	0.0	78
Biology	20	1.0	1.9	0.1	89

Source: Commission on Professionals in Science and Technology Professional Women and Minorities *A Total Human Resources Data Compendium* 16th edition (November 2006); www.cpst.org

At the annual meeting of the National Organization for the advancement of Black Chemists and Chemical Engineers (NOBCChE), recruiters from industry make contacts with promising graduates as early as the third year of graduate school. Students who continue in academia are the rare, passionate few, says Warner.

He and Velasco both say it is imperative that senior academic scientists do a better job of presenting the positive aspects of academia, including intellectual freedom and flexibility, and ultimately higher salaries and stability. The national meetings of NOBCChE and the Society for the Advancement of Chicanos and Native Americans in Science (SACNAS) give senior scientists the chance to mentor students across campus borders. They also fend off feelings of isolation and foster networking and professional development.

Diversity leaders say that if members of faculty search committees were to talent spot at these meetings like industry recruiters, they would see that, despite myths to the contrary, there are enough women or minority candidates to go around. NOBCChE meetings attract 600 students and SACNAS 1,000, with ever-increasing numbers of postdocs as well. Urry notes that some departments have been creative in attracting under-represented candidates in the same years that other departments claim there's a shortage.

"Some people understand how to do it by beating the bushes, being very aggressive, and paying attention to these biases," says Urry. "You are sitting on a *search* committee, not a sit-and-wait-for-it-to-come-over-the-transom committee." Academic departments should recruit at meetings, offer incentives to match industry, and let go of the sacrosanct 'open search' ideology that relies on job ads alone, says Urry. Departments should also consider how they can harness talent by employing the husbands and wives of staff members, catching available talent outside a full-scale search, or by doing broad-based searches and cluster hires of two or three female or minority candidates.

Urry says she often hears search-committee members say that they must hire a particular specialist because the department's students demand that expertise. "If your students are 25% minority and 50% female," she asks, "don't you think they demand professors who look like them?"

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

CLOSING THE GENDER GAP

Across Europe, women in science are typically outnumbered by men at every level. **Magdalena Wutte** explores how institutions, networking organizations and women themselves can help correct the imbalance.

According to the European Union (EU), this year is the 'European Year of Equal Opportunities for All'. This declaration, along with a slew of anti-discrimination legislation, suggests that the EU recognizes there's much to be done in the drive towards equality, as the numbers attest to. In the case of science, women remain under-represented, particularly at higher academic levels. And this disparity cannot simply be attributed to a lack of women pursuing science in the past: bias, it seems, remains.

The EU has not been short of initiatives to try to reverse this trend — and progress is being made. Several groups have been established to improve networking among women, but governments and lobby groups can only do so much. Academic institutions — and women themselves — have their part to play in increasing women's participation in the scientific arena.

"The situation has improved a lot in the past 20 years," says Daniela Corda, director for research and development of the Consorzio Mario Negri Sud, a major research institution in Santa Maria Imbaro, Italy. But, she adds, growing awareness and an absence of open discrimination are not sufficient to substantially increase the number of women in higher academic positions.

The most recent figures available suggest that Europe is still dealing with significant attrition by women after they've earned their PhDs (see Moving up or moving out). In 2006, the European Commission (EC) reported that although 40% of PhD students in the natural sciences are female, only 11.3% of the professor, research director and other top positions are occupied

by women[1]. In engineering and technology, 21.9% of PhD students are female, but this total dips to 5.8% at the highest levels of academia[1].

Furthermore, the average proportion of women on scientific boards is 24% (Norway and Finland, with 48% and 47%, respectively, stand in clear contrast to countries such as Italy and Poland, with 13% and 7%)[1]. And research funding also suggests a gender gap; in 17 of 26 European countries, men have higher success rates for securing funding[1].

Off-balance

The EU has been trying to address this imbalance. So far, its greatest success has been within its own organizations. The EC has almost reached its own target, set in 1999, of 40% for women on scientific boards and agencies[2]. According to Johannes Klumpers, head of unit for 'science culture and gender issues' at the EC's directorate for research, the numbers have increased from around 10% in 1998 to about 34% in 2006.

Europe-wide, the EU would like to increase the number of women in higher scientific ranks to 25%[3]. The EU first started to address the issue in 1999, when it set up an evaluation committee known as the 'Helsinki Group'. The sociologists and natural scientists on the panel, all hailing from EU member countries, drafted reports on the situation in their countries. The group also appointed 'statistical correspondents', based at national universities or private institutes, in order to ensure European statistics were comparable across countries. The reports of the Helsinki Group and their correspondents serve as guidelines for the EU and individual countries.

The Max Planck Institute's Mary Osborn recommends institution-wide changes to help recruit and support women scientists.

Mary Osborn, who is head of the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, and has chaired several panels on women in science thinks Europe would benefit from a programme similar to the US National Science Foundation's ADVANCE initiative. This programme provides funds to selected higher-education institutions to support institution-wide changes designed to increase participation and advancement of female scientists and engineers. These range from alteration of recruitment practice (such as actively campaigning for women when advertising jobs) to easing everyday life for scientists with children — for example, by creating day care centres or nurseries integrated into institutions.

Making contact

The EU's newest development is a central forum for exchange among women scientists across Europe, which had its first general assembly in April. The European Platform of Women Scientists (EPWS) is an umbrella organization for national women's networks and groups lobbying for women in science at EU level.

"It is a unique exchange forum" says Flavia Zucco, head of research at the Institute of Neurobiology and Molecular Medicine in Rome and member of EPWS's advisory board. EPWS meetings provide an opportunity to talk with national- and EU-level politicians, and to learn about the situation in other countries. "It is much easier to demand improvements such as flexible work shifts if you can point out that they are already standard in other European countries," Zucco says.

The National Contact Centre — Women in Science in Prague, an EPWS member, exemplifies how lobby work can function in practice. The EU-funded centre helps women whose position or work is suffering as a result of purported discrimination. Women can get advice on questions of labour law, and can also access legal and psychological help in more drastic cases of discrimination or sexual harassment.

By networking across institutes and national borders, women hope to penetrate and overcome the 'old-boys' networks'; established institutional structures that

Neurobiologist Gaia Tavosanis received valuable career advice from a mentoring project for women in science.

often make it difficult for them to penetrate the higher ranks. Women need to form their own connections early in their careers, emphasizes Gaia Tavosanis, head of a junior research group at the Max Planck Institute of Neurobiology in Munich, Germany. She says that the mentoring women network FemmeNet, run by the Max Planck Minerva Foundation, has helped her to plan her career.

A programme called 'ProFil' — run jointly by the Technical University, the Humboldt-University and the Free University of Berlin — goes further, aiming to help women who are preparing for professorships to advance. It offers professional training in skills such as mentoring, university administration, scientific presentation skills and grant writing. And at monthly dinners, ProFil attendees can make contact with politicians, industry representatives and journalists — helpful career-long contacts that could otherwise be difficult to establish.
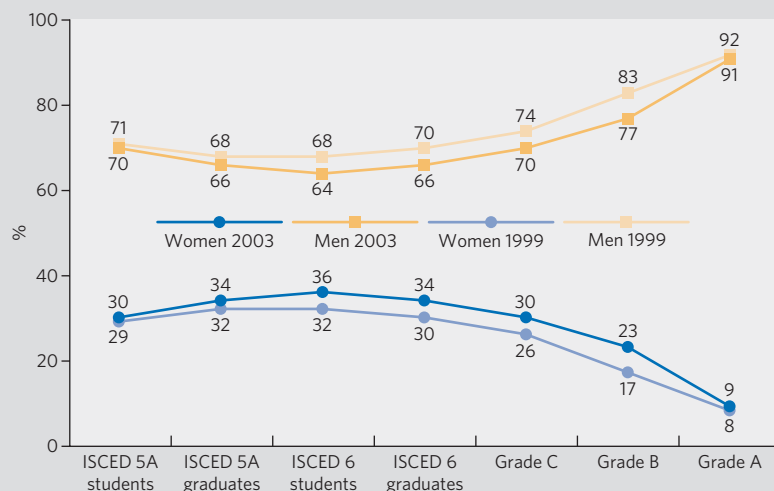
Caren Tischendorf, who was appointed professor of mathematics at the University of Cologne in 2006, especially endorses the motivation she gained from the programme. "Knowing that other women are in the same situation as you is very helpful," she says. "It strengthens your confidence that what you are trying to achieve is possible". The programme has been a huge success: of the 42 participants since 2004, 21 have been appointed full professors, and six junior professors. On the basis of this success, the programme was prolonged this year.

Progress towards equal opportunities for women has undoubtedly been made in recent decades, and particularly during the past few years. The challenge now is to make sure that initiatives to help women advance in scientific fields become the rule rather than the exception, whether through EU policies, institutional policies, or grassroots online efforts.
Magdalena Wutte is a former intern in *Nature*'s Munich office.

1. European Commission *She Figures 2006, Women and Science, Statistics and Indicators* (EUR22049) (Luxembourg, 2006).
2. European Commission Communication *"Women in science, mobilizing women to enrich European research"* (Luxembourg, 1999).
3. Council of the European Union *Council Conclusions on reinforcing human resources in science and technology in the European Research Area* (8194/05) (Brussels, 2005).

MOVING UP OR MOVING OUT
How the proportion of men and women varies through Europe's scientific ranks

Women 2003   Men 2003   Women 1999   Men 1999

	ISCED 5A students	ISCED 5A graduates	ISCED 6 students	ISCED 6 graduates	Grade C	Grade B	Grade A
Men 2003	71	68	68	70	74	83	92
Men 1999	70	66	64	66	70	77	91
Women 2003	30	34	36	34	30	23	9
Women 1999	29	32	32	30	26	17	8

Grade A represents the highest level research posts. Grade B identifies researchers in intermediate positions. Grade C is the first post for a new PhD graduate. ISCED 6 refers to students in programmes leading to a PhD or other advanced degree, and ISCED 5A refers to students of tertiary programmes that yield marketable skills for a research profession. Data from ref. 1.

Olympic talent

Fit...for nothing?

Richard A. Lovett

Michael Hood's quest for the 2068 Olympics began when he was six weeks old. Of course, it wasn't 2068 at the time. It was 2041, and Michael's parents were getting a routine assessment of his GeneChievement profile.

"Your son shows high athletic potential," the paediatrician said. "But he'll be even better if we fine-tune his gene expression a bit. What would you like him to be good at?"

"Not football," said his father, whose sports career had ended in a three-way collision between himself, the East Binghamton striker and a fan carrying something that appeared to be a tuba. "How about basketball?"

"Sorry, growth hormone is prohibited. I was thinking about muscles."

There are two types of fibres, the doctor explained, one for endurance, one for strength. "Unless we do something, he's going to have about half of each, which isn't ideal for much except football. You're better to shift it one way or the other. All it takes is a hormone shot. One way he's a sprinter or weightlifter, the other he's a long-distance type."

That was easy. Michael's father had fond memories of the days before the tuba, when he could run for ever and never get tired.

"Can you make him tough, too?"

"Sure." The doctor sneaked a peek at his palm computer and pulled up a folder labelled neurotransmitters. "How competitive would you like him to be?"

Time did its usual thing and the boy grew. He ran through woods and fields. He climbed trees and fell out of them because although he was whippet-thin and fast, feats of strength eluded him.

It was during one of his periodic stints in hospital that another doctor visited. "They didn't give you enough strength," the new doctor said. "But we have a technique for dealing with that. Actually, you're lucky. The sports authorities allow us to fix mistakes if people were improperly modified first time around. We can make you better than if they'd gotten it right back then. It's painless, but it uses a virus to reprogramme your muscle cells, so you might feel a bit under the weather for a few days."

Soon, Michael was both strong and fast — although at first that simply helped him find more sophisticated (or not-so-sophisticated) ways to show off. When he was 16, after a particularly bad brush with

gravity involving a flexboard and a blonde, yet another doctor visited, with an older man in tow.

"They certainly got the toughness part right," this doctor said. "Unfortunately, the same neurotransmitters can make you stupid."

"I get good marks!"

"Not that kind of stupid." The doctor gestured to his guest. "This is Coach Brooks. Sports are changing, largely thanks to gene-tweaked kids like you. He'll show you how to avoid spending so much time in here."

The 'how' turned out to be virtual reality. "There's no longer any need to risk getting hurt," the coach said. "I can set you up with a sensory-feedback suit in which you can be as competitive as you like without leaving home. Once you get used to it, you'll barely know the difference."

He turned to Michael's father. "It's perfectly safe. He could try triple flips on a flexboard and land on his head, and it would merely give him a bit of a thump. Not that I'm going to teach him flexboard tricks. He's always had running talent; if he's not hurt all the time, he can be really good. By the time he's in his prime, VR will be an Olympic sport."

Coach Brooks was wrong about one thing. By the time Michael was ready, virtual reality wasn't just part of the Olympics, it *was*

the Olympics. There was no need to build a stadium, no need to travel — although there was a bit of hubbub at first about the certification of VR stations to avoid secret… improvements.

Michael turned out to be as good as anyone could wish. He chose to double in the 5,000 and 10,000 metres, jostling elbows in the pack at the start, then pulling away with two others in what was ultimately a test of whose gene doctor had found the optimal mix of muscles and neurotransmitters.

Toughness triumphed and Michael, straining in his climate-controlled suit thousands of kilometres from his rivals, pulled through to double gold. He only wished he'd gone for the triple and added the marathon to the mix.

Then it was over. Coach Brooks came and took back the suit because Michael would be too old to make another go for gold four years later. Besides, there was a cadre of even better gene-crafted athletes coming up behind him.

He was famous, but nobody knew his face because the icon he'd worn in the games belonged to an even-more-famous musician — picked, of course, to be recognizable to the millions thronging the virtual grandstands.

One day, Michael found himself in a park, watching joggers. He'd bought a pair of shoes — his first real ones in years — so he laced them up and fell in behind a young woman who had programmed her shirt to display his Olympic icon crossing the finish.

Unfortunately, Coach Brooks had been wrong about another thing: VR wasn't quite the same as real running. Superbly trained for one, but not the other, Michael laboured just to keep the woman from pulling away.

But he still had his gene-tweaked toughness. And the woman was not only surprisingly good for a recreational runner, she was pretty. Michael's social skills were a bit rusty, but the shirt said that if he could catch her, she'd be easy to talk to. Pushing hard to catch up while trying to make it look easy, he wondered: if he and a natural athlete like her hit it off and found the right doctor, could their child win the triple? ■

Richard A. Lovett is a four-time winner of *Analog* **science-fiction magazine's AnLab reader's choice award. He also writes for** *New Scientist, National Geographic News* **and** *Running Times,* **and is a coach and co-author of two running books with marathon great Alberto Salazar.**

JACEY